# Single-Cell Analysis Reveals Protein Abundance-Dependent Signaling Network Modulations

**Dissertation**

**zur**

**Erlangung der naturwissenschaftlichen Doktorwürde**

**(Dr. sc. nat.)**

**vorgelegt der**

**Mathematisch-naturwissenschaftlichen Fakultät**

**der Universität Zürich**

**von**

Xiaokang LUN

**aus**

der V.R. China

**Promotionskommission**

Prof. Dr. Bernd Bodenmiller (Vorsitz und Leitung der Dissertation)

Prof. Dr. Lucas Pelkmans

Prof. Dr. Fabian Theis

Zürich, 2018

This work was carried out in the time between April 2012 and August 2017 in the laboratory of Prof. Dr. Bernd Bodenmiller at the Institute of Molecular Life Sciences, University of Zurich.

## Summary

Signaling network is central in controlling cellular processes including growth, proliferation, survival, differentiation, and responses to stress. Deregulation of signaling networks lead to uncontrolled cellular behaviors, causing the emergence of diseases, such as cancer. Independent of mutations which alter protein catalytic function, overexpression of signaling proteins has been shown to drive the development of many cancer types and identified as a key factor for drug resistance. However, the effects of overexpressed signaling proteins on the intracellular signaling networks have rarely been studied due to the facts that, first, cellular signaling heterogeneity and tumor genomic instability often leads to highly heterogeneous protein expression levels that signaling behaviors specific to each quantities of overexpression cannot be analyzed with conventional bulk measurement; and second, widely used single cell analysis profiles cell genomic and transcriptomic information that is known to poorly correlated with protein expression levels and cannot be used to predict protein phosphorylation. An approach that can measure signaling network behaviors in a multiplexed manner with single-cell resolution is a key to understand signaling protein abundance-modulated network states and dynamics. Mass cytometry is a recently developed technique that applies inductively coupled plasma time-of-flight mass spectrometry (ICP-TOF-MS) in combination with metal isotope-labeled antibodies to quantify protein or protein modifications at single-cell level. Yet mass cytometry enables high throughput analysis with over 40 proteomic parameters that has been shown to be the most suitable approach for single-cell signaling network analysis.

Using mass cytometry, in coupling with transient transfection which produces gradient expression levels for a GFP tagged protein of interest (POI), we established and thoroughly validated a method to quantify the network modulation induced by any overexpressed POI-GFP in an abundance-dependent manner. In a proof of principle analysis, we applied this method to 20 central signaling proteins in human embryonic kidney HEK293T cells to reveal that by altering signaling protein expression levels, signaling network states and the dynamics in responses to extracellular stimulation are modulated. Further, we used this method in a kinome- and phosphatome-wide screen for protein abundance-dependent effects. In total of 649 GFP-tagged POIs were individually overexpressed in HEK293T cells where signaling behaviors in the cancer-related signaling network were measured. We developed a statistical measure, called BP-$R^2$, to quantify the signaling relationship strength between an overexpressed POI and a measured phosphorylation site that enables a classification of kinases and phosphatases based on their abundance-dependent effects. Comparing to a literature-curated signaling interaction

database, we detected 208 pairs of the novel signaling relationships. The follow-up in-depth analyses with 1-hour EGF stimulation time course revealed oncogenic-like signaling behaviors that, first, ERK specific phosphatase overexpression prolonged the duration of proliferative signal; and second, tyrosine kinase overexpression induced ligand-independent ERK signaling activation. Later in melanoma A375 cells that express $BRAF^{V600E}$ mutant, we validated the abundance-dependent signaling relationship to p-ERK as a potential biomarker for the resistance to BRAF inhibition.

Importantly, we also found that mass cytometry-based single cell intracellular signaling analysis can be confounded by cell volume and cell cycle that usually biased computational analysis and resulted in unprecise data interpretation. Hence, we developed an experimental and computational approach called CellCycleTracer that can be used to correct for these confounding factors. Applying to data of TNF-a stimulated THP-1 cells, we showed CellCycleTracer improved the performance of follow-up analysis, especially for correlation- or mutual information-based network reconstruction.

## Zusammenfassung

Das Signalnetzwerk spielt eine zentrale Rolle bei der Steuerung zellulärer Prozesse, einschließlich Wachstum, Proliferation, Überleben, Differenzierung und Reaktionen auf Stress. Die Deregulierung von Signalnetzwerken führt zu unkontrolliertem zellulärem Verhalten, was zur Entstehung von Krankheiten wie Krebs führt. Unabhängig von Mutationen, die die katalytische Funktion des Proteins verändern, wurde gezeigt, dass die Überexpression von Signalproteinen die Entwicklung vieler Krebsarten vorantreibt und als Schlüsselfaktor für Arzneimittelresistenz identifiziert wurde. Die Auswirkungen von überexprimierten Signalproteinen auf die intrazellulären Signalnetzwerke sind jedoch selten untersucht worden, da die zelluläre Signalisierungsheterogenität und tumorgenomische Instabilität oft zu sehr heterogenen Proteinexpressionsniveaus führt, die Signalverhaltensweisen spezifisch für jede Überexpressionsmenge nicht können mit herkömmlicher Massenmessung analysiert werden; und zweitens weit verbreitete Einzelzellanalyse-Profile, genomische und transkriptomische Zellinformationen, von denen bekannt ist, dass sie schlecht mit Protein-Expressionsspiegeln korrelieren und nicht zur Vorhersage der Proteinphosphorylierung verwendet werden können. Ein Ansatz, der das Signalnetzwerkverhalten in einer multiplexen Art und Weise mit einer Einzelzellenauflösung messen kann, ist ein Schlüssel zum Verständnis von durch Signalprotonenhäufigkeiten modulierten Netzwerkzuständen und -dynamiken. Die Massenzytometrie ist eine kürzlich entwickelte Technik, die induktiv gekoppelte Plasma-Flugzeit-Massenspektrometrie (ICP-TOF-MS) in Kombination mit Metallisotop-markierten Antikörpern zur Quantifizierung von Protein- oder Proteinmodifikationen auf Einzelzellniveau anwendet. Die Massenzytometrie ermöglicht jedoch eine Hochdurchsatzanalyse mit über 40 proteomischen Parametern, die sich als der am besten geeignete Ansatz für die Einzelzellen-Signalnetzwerkanalyse erwiesen hat.

Unter Verwendung von Massenzytometrie, in Kopplung mit transienter Transfektion, die Gradientenexpressionsniveaus für ein GFP-markiertes Protein von Interesse (POI) erzeugt, haben wir eine Methode zur Quantifizierung der Netzwerkmodulation, induziert durch überexprimiertes POI-GFP, in abundanzabhängiger Weise etabliert und gründlich validiert . In einer "proof of principle" -Analyse haben wir diese Methode auf 20 zentrale Signalproteine in humanen embryonalen Nieren HEK293T-Zellen angewendet, um zu zeigen, dass durch Veränderung der Signalprotein-Expressionslevel die Signalnetzzustände und die Dynamik der extrazellulären Stimulation moduliert werden. Darüber hinaus verwendeten wir diese Methode in einem kinom- und phosphatomweiten Screening auf Protein-Häufigkeits-abhängige Effekte.

Insgesamt wurden 649 GFP-markierte POIs in HEK293T-Zellen individuell überexprimiert, wobei das Signalverhalten im krebsbezogenen Signalnetzwerk gemessen wurde. Wir haben ein statistisches Maß namens BP-R2 entwickelt, um die Signalstärke zwischen einem überexprimierten POI und einer gemessenen Phosphorylierungsstelle zu quantifizieren, die eine Klassifizierung von Kinasen und Phosphatasen basierend auf ihren abundanzabhängigen Effekten ermöglicht. Im Vergleich zu einer Literatur-kuratierten Signalinteraktionsdatenbank konnten wir 208 Paare der neuartigen Signalbeziehungen nachweisen. Die Nachuntersuchungen mit einem 1-stündigen EGF-Stimulationszeitverlauf zeigten onkogen-ähnliche Signalverhaltensweisen, die zunächst eine ERK-spezifische Phosphatase-Überexpression verlängerten die Dauer des proliferativen Signals; und zweitens induzierte Tyrosinkinase-Überexpression ligandenunabhängige ERK-Signalaktivierung. Später in Melanom-A375-Zellen, die BRAFV600E-Mutante exprimieren, validierten wir die Abundanz-abhängige Signalisierungsbeziehung zu p-ERK als einen potentiellen Biomarker für die Resistenz gegen BRAF-Hemmung.

Wir fanden auch heraus, dass die intrazelluläre Einzelzell-Signalanalyse auf der Basis der Massenzytometrie durch das Zellvolumen und den Zellzyklus verfälscht werden kann, was in der Regel zu einer ungenauen Dateninterpretation führte. Daher haben wir einen experimentellen und computergestützten Ansatz namens CellCycleTracer entwickelt, mit dem diese Störfaktoren korrigiert werden können. Bei der Anwendung von TNF-a-stimulierten THP-1-Zellen konnten wir zeigen, dass CellCycleTracer die Leistung der Nachuntersuchung, insbesondere für Korrelations- oder wechselseitige informationsbasierte Netzwerkrekonstruktion, verbesserte.

## Contents

# Introduction

## Signaling transduction in cell

Signaling is one of the central mechanisms that mediate cell communication and coordinate cellular functions, such as proliferation, differentiation, and energy metabolisms. The majority of signaling transduction pathways initiate at the plasma membrane (Groves and Kuriyan, 2010), where receptor proteins bind to their ligand signaling molecules, such as hormones and cytokines, or to surface proteins on an adjacent cell. In addition, signaling responses can be also triggered by changes of stress, ion concentration and extracellular environments (Clapham, 2007; Hetz and Saxena, 2017; Yu et al., 2015). Ultimately, intracellular signaling transductions result in cellular transcriptional modulations, phenotypical switches, and cell fate decision. We briefly summarize here essential signaling pathways that are most commonly studied in health and disease, grouped by their functions in cells and tissues:

### ***Proliferation, growth and survival***

Receptor tyrosine kinase (RTK) is a large group of signaling proteins that share the same molecular architecture and the conserved activation mechanism (Lemmon and Schlessinger, 2010). As key regulators of cell proliferation, growth, and survival, RTK signaling is commonly activated by the binding of a growth factor to the extracellular domain of RTK protein (Grecco et al., 2011). Then, series of downstream phosphorylation events occur consecutively, processing the signal transduction through pathways, including the PI3K/AKT/mTOR axis, the MAPK/ERK cascade, and the PKC pathways (Hennessy et al., 2005; Oliva et al., 2005; Steelman and Chappell, 2011), to rely signal from the activated RTKs into the nucleus where the induced transcriptional program happens (Schlessinger, 2000). In contrast to the RTK-related pathways, the Hippo pathway ceases cell proliferation and induces apoptosis via the inhibitory phosphorylation and ubiquitin-mediated degradation on transcriptional regulators YAP/TAZ, as a central mechanism controlling organ size and tissue homeostasis (Yu et al., 2015).

### ***Differentiation and homeostasis***

Several signaling pathways have been indicated to play central roles in stem cell differentiation, maintenance, and tissue development, including Wnt, Notch, Hedgehog (Hh) and the

transforming growth factor beta (TGF-β) signaling pathways. The canonical Wnt signaling pathway is essential in the embryonic development and the damage-induced tissue regeneration (Takebe et al., 2011). Activation mechanisms of Wnt signaling involve Wnt ligand binding with the heterodimeric receptor complex that is consisting of a Frizzled and an LRP5/6 protein. The subsequently phosphorylated LRP6 recruits the β-catenin destruction complex to the membrane. The complex then falls apart, disrupting the GSK3-mediated inhibitory phosphorylation on β-catenin, releasing free β-catenin molecules that are subsequently translocated into the nucleus to induce target gene transcription (Clevers and Nusse, 2012; Lien and Fuchs, 2014). The Hedgehog (Hh) signaling pathway functions in maintaining stem cell properties and controlling tissue polarity. Multiple hedgehog ligands (e.g., IHh, DHh, SHh) can activate the Hh signaling after binding to the transmembrane receptor patched (Ptch), allowing the activation of smoothened (SMO) and the release of Gli transcriptional factor into the nucleus (Takebe et al., 2011). Multiple phosphorylation sites on Gli proteins have been discovered that they determine the discrete states of Gli activity and tissue transcriptional landscape (Niewiadomski et al., 2014). Notch signaling regulates many processes in embryogenesis and tissue homeostasis. Activation of Notch signaling requires cell-to-cell contact that facilities ligand-receptor binding (e.g., Jagged1 to Notch1). The Notch receptor is then cleaved that enables Notch intracellular domain (NICD) to be transported into the nucleus for the transcriptional activation of target genes (Bray, 2016). TGF-β signaling is one of the most essential pathways in development that it is required in many developmental programs including differentiation, tissue hemostasis and regeneration. In the canonical TGF-β pathway, ligand binding induces the formation of a hetero-tetrameric receptor complex, from where the cytoplasmic Serine/Threonine kinase domain of TGF-β receptor is activated to phosphorylate downstream SMAD proteins. Upon the nuclear entry, transcription factors SMAD2/3 initiate the expression of TGF-β regulated genes (Massagué, 2012).

***Responses to stress***

MAPK branches, including the JNK and the p38 cascades, are primary responders to extracellular stress that they can be triggered by cytokine stimuli, such as tumor necrotic factor (TNF) and interleukin-1 (IL-1), and can be activated by environmental stresses, such as osmotic or heat shock, radiation and redox stress (Cuenda and Rousseau, 2007; Davis, 2000). Signaling dynamics on the JNK and p38 pathways determines cell fate: sustained activation of JNK and

p38 pathways leads to ROS-induced apoptosis, while their transient activation, on the contrary, promotes cell survival (Cuenda and Rousseau, 2007; Davis, 2000). Moreover, p53 senses the stress of DNA damage or the cellular hyperproliferation states, activating to induce transcription programs related to DNA repair mechanisms and to mediate growth arrest and apoptosis. The p53 pathways are crucial for innate tumor suppression given their key roles in the maintenance of genomic stability and controlled cell proliferation (Meek, 2009). Lastly, AMPK responds to the cellular nutrient stress and maintains energetic homeostasis by regulating multiple downstream pathways, including the inhibition of AKT/mTOR-mediated cell growth (Mihaylova and Shaw, 2011).

***Immune cell signaling***

Regulation of the immune response is a complex process that involves the coordination of multiple ubiquitous or cell type-specific signaling pathways. In brief, the innate immune response initiates with the recognition of pathogen-associated molecular patterns (PAMPs) by the pattern recognition receptors (PRRs) expressed on the surface of many cell types, including macrophages, dendritic cells, monocytes and neutrophils (Newton and Dixit, 2012). As the major PRRs, toll-like receptors (TLRs), upon pathogenic ligand-engagement, activate the IKK/NKκB pathway and MAPK cascades (ERK, JNK and p38 branches) that lead to the expression of cytokines and pro-inflammatory mediators (Banerjee and Gerondakis, 2007). Meanwhile, the antigens can be recognized by the B cell receptor complex consisting of immunoglobulin (IgG) and CD79α/β that will trigger formation of the signalosome where kinases, including SYK, BTK and LYN can be phosphorylated. Subsequently, downstream signaling mediated by PI3K/AKT, PLCγ/PKC and NFκB pathways are activated to induce the proliferation of antigen-specific B cells (Woyach et al., 2012). Importantly for the activation of adaptive immune response, T cell receptor (TCR) is initiated after the recognition of cognate peptide–MHC molecules. Similarly, signaling pathways of PI3K/AKT, PLCγ/PKC/NFκB axis and the MAPK cascades are activated to induce transcriptional programs regulating T cell activation, differentiation and the cytolytic activity (Brownlie and Zamoyska, 2013; Pennock et al., 2013). As a crucial part in the immune regulation, cytokine signaling pathways commonly activates the JAK/STAT kinases that rapidly transduces receptor signals into the nucleus, modulating transcriptional events that are specific to the cell types and stimuli (Bezbradica and Medzhitov, 2009).

## Signal computing as networks

### ***Pathway crosstalk***

Each of above pathways engages specific signaling transduction processes to determine transcriptional responses. However, signaling pathways are rarely independent in living organisms; rather, they are bridged by inter-pathway protein interactions, named crosstalk. In signaling response, diverged pathways can be stimulated by the same kinase or lipid responder, while pathways might also couple to combinatorically control one signaling output (Cohen and Frame, 2001; Corbalán-García and Gómez-Fernández, 2006; Nishi et al., 2015). For example, as a central coordinator between various signaling pathways, GSK3 can be directly regulated by a broad range of upstream signaling proteins, including AKT, mTOR, AMPK and PKC (Cohen and Frame, 2001), and can regulate the functional state of about 100 reported substrates (Sutherland, 2011). TGF-β signaling, besides the canonical pathways depending on the activation of SMAD2/3 as described above, also induces multiple non-canonical pathways by crosstalk that transduces signals into MAPK cascades, PI3K/AKT pathways and cell cycle regulating proteins (Derynck et al., 2014). Interestingly, the ERK-dependent SMAD2/3 activation and nuclear translocation has also been revealed and the inhibition of MEK can disrupt the SMAD function (Blanchette et al., 2001; Hayashida et al., 2003). Moreover, as most intensively studied, multiple classic crosstalk presents between the MAPK/ERK pathways and the PI3K/AKT/mTOR axis, co-regulating the cellular functions of proliferation, growth and survival (Steelman and Chappell, 2011). MAPK/ERK signaling can also crosstalk to the JAK/STAT pathway and the Hippo pathway (Rawlings et al., 2004; Romano et al., 2014). Indirectly, the pathway crosstalk can also be achieved via phosphatases which are often functionally unspecific and participate in the dephosphorylation of multiple substrates (Jeffrey et al., 2007).

### ***Study cell signaling in networks***

Signaling pathways, with the crosstalk connections, joint to form complex signaling networks that integrate several layers of signaling events and coordinate to compute the information from signaling inputs, such as extracellular stimuli, into transcriptional changes that determine cell phenotype and cell fate. The activity and dynamics of each specific signaling node is therefore multi-parametrically dependent within the network structure, which furthers the complexity in

studies of cell signaling behaviors. Critically, the primary cue to decipher the networked regulation of signaling response is to understand the signaling protein connectivity. To this end, studies implement biochemistry assays, including yeast-two-hybridization (Y2H), immunoprecipitation (IP) and luminescence reporter-based techniques, in combination with proteomics analysis at the systems level and computational network modelling and inference to promote a comprehensive and precise mapping of the protein interactome in global level and for particularly signaling pathways (Barrios-Rodiles et al., 2005; Breitkreutz et al., 2010; Couzens et al., 2013; Horn et al., 2011; Hughey et al., 2009; Linding et al., 2007; Tewari et al., 2004). Meanwhile, databases of protein interaction networks have been created and developed that largely promote the information accessibility and facilitate the researches applying network-based prediction and modeling of signaling behaviors (Linding et al., 2007; Saez-Rodriguez et al., 2011). The STRING database (Szklarczyk et al., 2017) integrates diverse sources of data and information for protein physical interactions, as well as the known and predicted protein functional associations, allowing querying of protein connectivity and network-dependent functional enrichment analysis. Complementarily, the OmniPath (Türei et al., 2016) database assembles 34 literature-curated databases (e.g., SIGNOR (Perfetto et al., 2016) and PhosphoSite (Hornbeck et al., 2015)) for protein interactions that drastically enhance the network coverage and interaction confidence.

**Heterogeneity in cellular signaling network**

Achievements in protein interactome and network mapping have greatly contributed in understanding the biochemical processes of cellular signaling responses. However, for every defined cell types and under specific conditions, the network responses can be highly variable due to many sources of genetic and non-genetic heterogeneities.

***Signaling protein abundance***

One of the major determinants of network behaviors is the signaling protein abundance. A signaling network eventually assembles numerous biochemical reactions (e.g., phosphorylation and dephosphorylation), in which the substrate concentration determines the reaction rate (Heinrich et al., 2002). By regulating the presence and expression level of signaling proteins, cells modulate the kinetics of their signaling transduction processes and the network dynamics,

ultimately determining outputs (both transcriptomic and proteomic levels) in response to the same stimulation (Lun et al., 2017). Aberrant expression levels of signaling proteins are often disease-related and can be attributed to genetic or epigenetic reasons. The major genetic cause is chromosomal abnormalities, including chromosomal rearrangement, amplification and deletion, and the generation of neochromosomes (Garsed et al., 2014; Holland and Cleveland, 2012), leading to gene copy number alteration (CNA) of signaling mediators, such as kinases and phosphatases, which has been identified in many cancer types (Fleuren et al., 2016; Julien et al., 2011). In humans, the epigenetic alterations are progressed with aging (Gonzalo, 2010) and can be induced in diseased-conditions, such as chronic inflammation that drives cancer progressions (Abu-Remaileh et al., 2015; Jeong et al., 2014). Additionally, deregulated transcription or translation of signaling proteins also leads to their abnormal expression levels, and can also be tumorigenesis (Bywater et al., 2013; Silvera et al., 2010).

***Mutation***

Protein mutants often alter the catalytic function of signaling nodes, leading to the re-shaping of signaling network structure and dynamics. Oncogenes, such as RAF, MEK and PI3K, and tumor suppressors, including PTEN and p53, once mutated, can drive the uncontrolled cell proliferation and growth, ultimately trigger cancer initiation (Mendoza et al., 2011). Six types of network-attacking mutations have been recently identified in a kinome-wide study with underlying signaling mechanisms (e.g., constitutive activation, network rewiring and genesis of phosphorylation sites) been categorized (Creixell et al., 2015). Importantly, due to the genomic instability, cancer progresses with accumulated and varied mutations (Abbas et al., 2013) that lead to discrete abnormities of signaling network behaviors at the single-cell level, which further complexes the network heterogeneity and can lead to treatment failures.

***Subcellular localization***

Signaling network responses even in cells with the same genetic background can also be highly variable. By altering the subcellular localization of signaling proteins, cells differ in their information processing in response to exogenous cues, contributing to the heterogeneity of cellular signaling network properties. For example, many known determinants, such as spatial constrain and lipid raft composition, influence density of receptor proteins on the plasma

membrane, resulting in differential downstream signal activities (Frechin et al., 2015; Grecco et al., 2011). Cells also modulate the dynamics of intracellular vesicular trafficking to intrinsically regulate protein localizations as a result of varied phenotypical states and signaling contexts, giving rise to heterogeneous signaling behaviors (Goh and Sorkin, 2013; Stallaert et al., 2017; Stenmark, 2009). Nuclear translocation is an important regulatory process for many cytosolic or surface signaling proteins functioning as transcriptional factors or co-regulators (Rawlings et al., 2004; Steelman and Chappell, 2011; Takebe et al., 2011). Dynamics of nuclear translocation are often varied between cells (Regot et al., 2014), and can be highly oscillated (Kellogg and Tay, 2015), leading to varied signaling activities in cells. Interestingly, a recent study has discovered cytosolic compartmentalization generated from stress-induced liquid unmixing largely alter the intracellular concentration of functional kinases, thereby further complexing the signaling network behaviors and contributing to the network heterogeneity (Wippich et al., 2013).

***Microenvironment factors***

Microenvironment factors, including extracellular signaling ligands and matrix stiffness, cell-cell contacts and local crowding, are the extrinsic determinants for heterogeneous signaling network states and dynamics. Differential ligand binding determines the consistency of ERK signaling in PC-12 cells (a pheochromocytoma cell line derived from rat adrenal grand): treatment with EGF transiently activates the RAF/MEK/ERK cascade, resulting in cell proliferation; while NGF induces PKC activity which sustains the ERK signal, leading to differentiation (Santos et al., 2007). The stiffness of extracellular matrix has been shown to influence cellular signaling responses in the human breast MCF-10A cell line due to the altered integrin conformation and the subsequent downstream PI3K pathway modulations (Chaudhuri et al., 2014). Cell also senses the surrounding neighbors that several contact inhibition mechanisms exist to impact signaling network behaviors (Kim et al., 2011; Stallaert et al., 2017). In addition, local crowding has been indicated to activate focal adhesion kinase (FAK) that results in the alteration in membrane homeostasis and the signaling responses in the lipid-dependent PI3K/AKT signaling pathways, influencing cell growth and proliferation (Frechin et al., 2015).

Many other cellular variables, including the cell cycle stage (Rapsomaniki et al., 2018), also leads to differed signaling network states among single cells. This heterogeneity is crucial for biological process, such as cell differentiation, tissue development (Donati and Watt, 2015) and the maintenance of functional systems (Potente and Mäkinen, 2017). Nevertheless, for many

diseases including cancer, signaling heterogeneity also causes the increase of phenotypical complexity that may challenge the development of therapeutic interventions and reduce the treatment efficiency of a given drug intervention (McGranahan and Swanton, 2017). Thus, to understand the signaling network at single cell level, the causes of signaling heterogeneity and the related downstream effects on cellular transcription and phenotypical changes it is crucial and of great beneficial for health and disease.

**The single-cell era of signaling network analysis**

Single cell analysis has been performed since the invention of microscope. Conventional microscopic methods can be used to quantify protein abundance and study cellular and subcellular spatial properties, helped by genetic or immunological fluorescent protein tagging methods (Waters, 2009). Fluorescence resonance energy transfer (FRET) and proximity ligation assay (PLA) enable the application of microscopy in analyzing transient events during the dynamics process of signaling transduction and measuring stable interactions of signaling proteins at single-cell resolution (Ryu et al., 2015; Sundqvist et al., 2013). However, these methods are often limited by the multiplexing capacity that individual experiments using batches of different cells are required to measure series of variables, resulting in the loss of relationship between assessed markers and the difficulties to study signaling process mechanisms at the network level. The era of omics has made it possible to measure simultaneously the transcriptomic and proteomic information (Horgan and Kenny, 2011). Protein phosphorylation, one of the most critical post-translational modifications (PTMs) for signaling transduction can be globally analyzed with phosphoproteomic approaches (Riley and Coon, 2016). Yet these methods are not sensitive enough to be applied in single cell measurement. Developing approaches that combine multiplexed measurement with single cell-based analysis has been for decades the challenge for single-cell signaling network analysis. By multiplexing simultaneous antibody detection capacity in immune-single cell analysis (Angelo et al., 2014; Bandura et al., 2009; Gerdes et al., 2013; Giesen et al., 2014; Jungmann et al., 2014; Lin et al., 2015) and applying single cell signal amplification to increase sensitivities for omics methods (Chen et al., 2015; Ramsköld et al., 2012; Tang et al., 2009), recent innovations now allow exploring cellular phosphorylation landscapes or signaling regulatory network structure cell by cell in heterogeneous samples.

## ***Non-spatial immuno-single cell analysis***

### *Flow cytometry*

As one of the earliest attempt of cell single-cell signaling network analysis, flow cytometry uses fluorophore-labeled antibodies to detect and quantify the protein phosphorylation abundance. It has been applied in measuring states and relationships between multiple phosphorylation sites, as well as the correlation between phosphorylation states, cell functional readouts, and lineage-specific markers in complex populations (Perez and Nolan, 2002). Later, with the capability to measure 11 phosphoproteins and phospholipids simultaneously, flow cytometry-based single cell analysis are combined with inhibitor perturbation assays that enables the inference of signaling circuits and the reconstruction of signaling networks (Sachs et al., 2005). The development of fluorescent cell barcoding has greatly increased the throughput of flow cytometry-based intracellular signaling analysis which now can be implemented as a screening tool to quantify cellular response to kinase inhibitors for each individual cell type in heterogeneous populations (Krutzik and Nolan, 2006; Krutzik et al., 2008). However, due to the fluorescent spectrum overlapping, the multiplexity of flow cytometry remains limited that signaling network can only be sparsely or partially measured.

### *Mass cytometry*

Mass cytometry is recent developed based on an inductively coupled plasma time-of-flight mass spectrometry (ICP-TOF-MS) and a single cell sample introduction system (Bandura et al., 2009). Mass cytometry-based measurement applies metal isotope-tagged antibodies to label proteins or protein modifications in cells that highly increase the multiplexity. During the measurement, each stained single cell is vaporized, atomized, and ionized before the subsequent formation of an ion cloud that carries information of all applied antibodies that can be quantitatively analyzed by the TOF and recorded as high-dimensional single-cell proteomic readout (Bandura et al., 2009; Bendall et al., 2011). Mass cytometry is ready to quantify up to 50 intracellular phosphorylation sites simultaneously with high analytical throughput that can be stable around 500 cells per second and millions of events per sample. A mass-tag barcoding strategy allows simultaneously measuring hundreds of samples, eliminating batch effects that confounds conventional measurements, and largely reducing the workloads (Bodenmiller et al., 2012; Zivanovic et al., 2013; Zunder et al., 2015). Although minor spill-over between channels are found in mass cytometry data due to metal impurity, mass overlap, and oxidation(Bendall et al.,

2012), they are manageable with proper experimental design and can be removed computationally (Chevrier et al., 2017a).

Mass cytometry has been used in drug screening for high-dimensional signaling (phosphorylation) responses (Bodenmiller et al., 2012). Relationships between all pairs of measured phosphorylation sites can be computed to infer network responses to a stimulus (Krishnaswamy et al., 2014) or to trace the network reshaping process through a biological time of phenotypical transition (Krishnaswamy et al., 2017). Coupling to the transient overexpression technique, mass cytometry based signaling profiling has been able to assess the protein abundance dependency on intracellular signaling network states and dynamics, revealing novel signaling mechanisms in overexpression-related cancer progression and drug resistance (Lun et al., 2017).

*Single cell immuno-sequencing*

Although the multiplexity of immune-based single cell analysis has been largely increased by mass cytometry, the limited number of available and assessable metal isotopes is still preventing the deep profiling of phosphoprotein network with broad coverage of phosphorylation sites. Nevertheless, two recent developed techniques, termed RNA expression and protein sequencing assay (REAP-seq) and cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), barcode antibodies with oligonucleotides that allows detecting targeted proteins in single cell sequencing and simultaneously quantifying RNA transcriptomes in the very same cells (Peterson et al., 2017; Stoeckius et al., 2017). More than one million distinct barcodes can be generated with a 12-bp oligo ($4^{12}$), making the measurable parameters in these methods unlimited. REAP-seq and CITE-seq are ready for single cell phosphoprotein network analysis to generate more comprehensive global signaling profiling.

*Lab-on-chip and microfluidics*

With the lab-on-chip technologies, single-cell barcode chip (SCBC) and single-cell western blotting (scWesterns) have been developed. Compared to the above cytometric methods, SCBC and scWesterns are more sensitive in detecting low abundant targets and can be applied to assess samples with fewer cells (Hughes et al., 2014; Shi et al., 2012; Wei et al., 2016). These approaches have been applied to resolve singe cell signaling network variations and the functional heterogeneity (Shi et al., 2012; Wei et al., 2016). Meanwhile, investigations of single

cell signaling kinetics largely benefits from microfluidic systems which allow fine time resolution and accurate dose control for a to-be-profiled stimulus (Ng et al., 2015).

## ***Non-spatial omics single cell analysis***

Immunostaining-based techniques allow multi-dimensional deep profiling of signaling networks at single-cell resolution, but also face three main limitations: first, the selection of measured targets are often based on prior knowledge, making these methods not exploratory; second, not all targets of interest are measurable due to the high dependency on antibody availability; third, quantifications cannot be performed across different antibodies, given their varied antigen binding affinities. Nevertheless, the development of several antibody-free omics studies have provided complementary techniques void of the above limitations.

### *Single cell proteomics by mass spectrometry*

A big challenge for single cell mass spectrometry is the relatively low sensitivity of the technique, partly due to the sample loss during processing. Often, protein extracts from single cells are not detectable. One recent study has applied tandem mass tags (TMT) labeling to embed single mammalian cells in hundreds of carrier cells to manage the sample loss (Budnik et al., 2018). This method measures proteome at single-cell resolution and can be furthered developed as single-cell phosphoproteomics for signaling network profiling.

### *Single cell transcriptomics and epigenomics*

Single cell sequencing techniques (Buenrostro et al., 2015; Tang et al., 2009) do not directly measure the protein abundance and functional modifications that are typically used in signaling network analysis. However, with the strength to quantify the global RNA expression and identify whole-genome transcriptional regulation landscapes, these approaches infer the transcriptional regulatory network and the dynamics of signaling pathways in response to a stimulus. For example, single cell RNA-seq and has been used in uncovering a paracrine signaling-required repression of inflammatory program (Shalek et al., 2014). Single-cell epigenome can now be measured with the assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Buenrostro et al., 2013; Cusanovich et al., 2015). Coupling single-cell transcriptomics and epigenomics analyses provides a powerful approach to profile the network of transcriptional regulation during stem cell differentiation and allows identification of curtail signaling pathways

during the transition states from quiescence to proliferation and differentiation (Buenrostro et al., 2017; Guo et al., 2017).

## ***Spatial immuno-single cell analysis***

Spatial variables (e.g., cell contacts and protein localizations) might act as crucial determinants during the processing of cellular signaling information. These properties cannot be assessed with the above non-spatial single cell analytical methods as cell detachment is required for sample acquisition. Imaging-based cytometry, proteomic and transcriptomic techniques can preserve the cellular spatial information and are also capable to resolve in subcellular details for protein localization. The additional spatial dimension gained with these approaches is often a cue to understanding the source of cellular heterogeneity that helps the profiling and prediction of signaling network behaviors.

### *Microscopic sequential imaging approaches*

Spatial information of protein localization and cell/tissue organization can be acquired through fluorescence microscopic measurements of cell monolayers or tissue sections. However, the fluorescence spectrum overlap limits detectable channels that can be detected in a simultaneous measurement. To achieve the high multiplexity required for signaling network profiling, technologies have been developed the very same specimen can be sequentially imaged for serial targets of interest without influencing the antigen abundance and disrupting the tissue structure. The first set of current sequential imaging approaches applies fluorophore-labeled antibodies to detect targets of interest (Gerdes et al., 2013; Lin et al., 2015; Schubert et al., 2006). Specifically, multi-epitope-ligand cartography (MELC) implements photo-bleaching after each round of antibody staining and imaging circle to remove the remained fluorescence (Schubert et al., 2006). Alkaline oxidation chemistry has been used in a recent developed method, called multiplexed Immunofluorescence (MxIF), to chemically inactivate the fluorescent dyes after imaging (Gerdes et al., 2013; Li et al., 2014). Cyclic immunofluorescence (CycIF) combines the oxidative inactivation and enzymatic antibody cleavage for sequential imaging(Lin et al., 2015). The second set of approaches apply DNA-labeled antibodies (Goltsev et al., 2017; Jungmann et al., 2014; Wang et al., 2017). Unlike the above methods which require rounds of antibody staining and are usually time-consuming, DNA-labeled antibodies are simultaneously stained on the specimen. During the detection process, sequence of the DNA oligo serves as barcode for each antibody that it can be detected with fluorescent probes in Exchange-PAINT

(Jungmann et al., 2014) and DNA-Exchange-Imaging (DEI) (Wang et al., 2017), or with fluorophore-labelled dNTPs (Goltsev et al., 2017) in CO-Detection by indEXing (CODEX) (Goltsev et al., 2017). These approaches allow profiling the spatial signaling heterogeneity and the understanding of tissue organization-related network variations (Gerdes et al., 2013; Graf and Zavodszky, 2017). Multiplexed super resolution imaging also allows three-dimensional (3D) spatial profiling and can be further applied to analyze intercellular and subcellular signaling events in 3D (Jungmann et al., 2014; Wang et al., 2017).

*Mass spectrometry-based imaging approaches*

Imaging mass cytometry (IMC) implements laser ablation (LS) as the core of sampling platform that antibody-stained tissue samples can be ablated spot by spot. A mixed argon and helium stream then transports the ablated materials successively into a mass cytometer where proteins and protein modifications, such as phosphorylation, can be quantified and subcellular level ($1\mu m^2$) spatial information is preserved (Bodenmiller, 2016; Giesen et al., 2014). With a recent improvement, IMC can analyze protein (including phosphoproteins) and RNAs in the same sample that enables, for example, research for understanding the transcriptional control under specific cellular and spatial signaling properties (Schulz et al., 2017).

Similarly to IMC, multiplexed ion beam imaging (MIBI) also applies metal-labelled antibodies in tissue staining. MIBI then uses oxygen duoplasmatron primary ion beam to liberate the antibodies that generate the secondary ion beam. Subsequently, a magnetic sector mass spectrometer detects the isotope abundance from the second ion beam from every pixel of analyzed sample (Angelo et al., 2014).

## ***Spatial omics in single cell analysis***

Current technologies have greatly increased the resolution and sensitivity of proteomics and transcriptomics that single-cell level spatial analysis by these approaches can be achieved.

*MALDI-imaging*

Matrix-assisted laser desorption/ionization (MALDI) imaging mass spectrometry can be used to globally detect bio-molecules, including lipids, metabolites, peptides and proteins (Schwamborn and Caprioli, 2010). Although MALDI-imaging is mainly used at tissue-level resolution, it allows unbiased quantitative and spatial profiling of signal-mediating lipidome and metabolome

(Sugiura et al., 2015), and can be further used in systemic assessments of disease states or drug responses (Nielsen et al., 2016; Schwamborn and Caprioli, 2010). Yet, novel MALDI-imaging platform has been developed with increased ionization efficiency, which pushes the resolution to a subcellular level of 5 [illegible] resolution (Soltwisch et al., 2015).

*Spatial transcriptomics*

Several spatial transcriptomics approaches have been established based on various techniques including fluorescent *in situ* sequencing (FISSEQ) (Lee et al., 2014), multiplexed error-robust FISH (MERFISH) (Chen et al., 2015) and spatial barcoding (Ståhl et al., 2016). These approaches are ready to, from the transcriptomic level, interpret signaling pathway activation states, and correlate such cellular variances to the cell-to-cell communication. Spatial transcriptomics are also powerful in understanding the remote cell signaling control mechanisms since mRNAs are used as expression readouts for secreted ligands (e.g., cytokines and chemokines) that are otherwise difficult to detect in proteome-based analysis (Schulz et al., 2017).

***Live cell imaging***

It is important to note that cell signaling transduction is a dynamic process that it cannot be fully understood from snapshot measurement of transient network states. Information of the time dimension, in addition to the multiplexed signaling profiling, is therefore necessary to systematically decode the causality of signaling behaviors and to characterize network kinetics (Koseska and Bastiaens, 2017). However, signaling events are mainly present intracellularly. They can be detected only after a fixation and permeabilization procedure that disrupts the signaling dynamics through time. Conventionally, serial snapshot information needs to be acquired to enable the rebuilding of time dimension and the computational reconstruction of signaling trajectories (Lun et al., 2017; Zi et al., 2011). Technically, these approaches cannot fully resolve to the transient events of signaling processing and the computation inference becomes complicated when measured signaling behaviors are highly heterogeneous. Several live cell imaging methods exploit the changes of protein physical properties (e.g., subcellular localization and proximity and modification) to monitor signaling events through time (Aoki et al., 2013; Burack and Shaw, 2005; Kellogg and Tay, 2015; Kuchenov et al., 2016; Lee et al., 2009;

Regot et al., 2014; Ryu et al., 2015). Although these methods are not highly-multiplexed yet, capturing central signaling node through time allows tracing the pathway and network behaviors.

*FRET*

Fluorescence resonance energy transfer (FRET) technology is developed based on the energy transfer from two proximate fluorophores, leading to the shift of the emission spectrum that can be captured by microscope. FRET can be used in monitoring the proximity of interactive signaling proteins (Burack and Shaw, 2005) or as bio-sensors for phosphorylation events to indicate the pathway activity in real time (Aoki et al., 2013; Ryu et al., 2015). These studies characterize single-cell temporal signaling states and correlate them with functional readouts, such as proliferation and differentiation. Given the broad fluorescent spectrum occupancy from each FRET sensor, multiplexing FRET experiments to study more complexed signaling network behaviors can be challenging. Several approaches to increase FRET multiplexity have been developed that by expanding selectable fluorophores or promoting the image decoding and error propagation scheme, up to six protein interaction/phosphorylation events can be measured simultaneously in a multiplexed FRET setup (Bunt and Wouters, 2017; Geißler et al., 2013; Hoppe et al., 2013). Alternatively, FRET biosensors can be used in combination with a multi-parameter imaging platform (FMIP) that the activity of 40 signaling proteins are imaged in individual cells to comprehensively infer network dynamics (Kuchenov et al., 2016).

*Kinase activity reporter*

Many kinases, such as ERK and NF □B, have nuclear translocation property in their active forms. Thus, they have been used in tracing signaling activities in real-time (Kellogg and Tay, 2015; Lee et al., 2009; Lidke et al., 2010). Studies applies single-cell kinase nuclear translocation methods have demonstrated the cellular heterogeneity in signaling dynamics (Lee et al., 2009) and noise-facilitated transcription output (Kellogg and Tay, 2015). A recently established system applies kinase translocation reporters (KTRs) to monitor the activity of key signaling mediators including JNK, p38 and ERK simultaneously, and uncovers temporal signaling response between the pathway crosstalk (Regot et al., 2014). An important strength of these live cell imaging technologies is the preservation of cellular natural states. The same imaged samples can be re-analyzed in other compatible single cell methods. For instance, a study has couple NFκB nuclear translocation analysis with single cell RNA-sequencing that reveals three distinct cell subpopulations with differential NFκB dynamic-related transcription profiles (Lane et al., 2017).

**Computational method for signaling network analysis using single cell information**

Advanced single-cell experimental approaches share one important common fact of generating high-dimensional multivariable datasets which bring challenges for conventional computation tools in analyzing signaling network properties. Nevertheless, multiplexed measurements allow systematically assess network states and dynamics in one single experiment in which the multivariate dependence and high-dimensional distribution are precisely preserved. Recent developed computational approaches apply statistical inference to reconstruct signaling network structure (Chan et al., 2017; Huang et al., 2016; Krishnaswamy et al., 2014; Sachs et al., 2005) and use mechanistic model to characterize network dynamics (Hasenauer et al., 2014; Loos et al., 2017).

For the reconstruction of signaling network, Bayesian model has been applied in flow cytometry measurement of 11 intracellular phosphorylation sites with individual treatments of nine small molecule inhibitors. Exploiting the nature cellular variances and the re-shaping of multivariate distributions upon perturbations, a probabilistic network can be assembled that it replicates known pathway relationships and predicts novel network causalities (Sachs et al., 2005). Alternatively, correlation based statistics quantify the relationship and dependency between measured parameters, and are therefore widely used to assess strength of signaling circuits and infer network structure and dynamics in both flow cytometry and transcriptomics data (Redell et al., 2013; Yuan et al., 2011). However, in complex signaling regulatory network, relationships between a pair of signaling proteins are often multi-parametric dependent and non-monotonic in shape. Correlation analysis often fails to reflect the true strength of these relationships. Based on the information theory, mutual information (MI) and maximal information coefficient (MIC) measure relationship between two variables independent of their linearity and continuity, allowing quantification of diverse and complex relationships (Kraskov et al., 2004; Reshef et al., 2011). Following the same line, a more advanced measure termed conditional Density Resampled Estimate of Mutual Information (DREMI) has been recently developed to quantify the mutual information in a density independent manner that removes the bias of cell distribution along the measured variables (Krishnaswamy et al., 2014). Network reconstructed and quantified by DREMI recapitulates well-known signaling processes. In combination with methods to trace biological time in a cell transition process (Bendall et al., 2014), DREMI is able to study the signaling network reprogramming during cellular phenotypical shifts (Krishnaswamy

et al., 2017). Another density independent measure, binned pseudo-R2 (BP-$R^2$), applies the classical $R^2$ statistics that is shown to precisely reflects strength of signaling relationships in steady states and dynamics (Lun et al., 2017).

As complementary methods, mechanistic models reveal biochemical insights of a given signaling network and the functional heterogeneity within a cell population. Ordinary differential equation (ODE) is commonly applied in such modeling that it uses mass action kinetics to determine the concentration of signaling nodes over time. ODE model is capable to study important network features, such as feedback loops (Hughey et al., 2009). A pilot single cell analysis has used ODE constrained mixture modeling to study the variability of phosphorylated ERK in the PC12 cell line responding to the NGF stimulation and mechanistically described two cell subpopulations with differential signaling responses caused by varied receptor abundance (Hasenauer et al., 2014). Further exploiting the cell variabilities, a hierarchal population model has been developed, in combination with the single cell modeling, to explain multiple levels of heterogeneity in NGF-treated PC12 cells simultaneously (Loos et al., 2017).

**Accounting for confounding factors**

Single-cell technologies have enabled characterization of differential signaling behaviors in a heterogeneous cell population that are often masked in conventional batch measurements. However, these advantages also come along with the challenge that multiple levels of confounding factors, including cell cycle and cell size, can bias the single-cell readouts (Behbehani et al., 2012; Buettner et al., 2015; Kowalczyk et al., 2015; Marguerat and Bähler, 2012; McDavid et al., 2016; Padovan-Merhar et al., 2015; Rapsomaniki et al., 2018) and need to be appropriately treated in the single-cell analysis.

One of the most important biological confounding factors is the cell cycle, as cells regulate differential signaling and transcriptional programs through the cell cycle phases to control cellular events, such as protein synthesis and DNA replication. Particularly in the cell signaling level, phosphorylation of signaling proteins have been shown to involve in the cell cycle regulation and vary through the cell cycle progression (Fisher et al., 2012; Gut et al., 2015). For single cell analysis, it is essential to distinguish the cell varieties caused by the cell cycle progression from other sources of heterogeneity. Multiple computational methods are now available to account for the cell cycle effects in single cell transcriptomes data, mass cytometry-

based single cell proteomics data and microscopic imaging analysis (Behbehani et al., 2012; Buettner et al., 2015; Gut et al., 2015; Rapsomaniki et al., 2018).

Single cell measurements typically use the abundance readouts of measured targets. However, as signaling network is an integration of biochemical reactions, the rate of signaling transduction event is determined by the substrate concentration (Hasenauer et al., 2014; Heinrich et al., 2002) which cannot be directly interpreted form the abundance measurement. Cell size has been demonstrated in multiple studies to confound single-cell abundance measurement that it linearly correlates with most of the measured mRNA or protein (modification) levels (Padovan-Merhar et al., 2015; Rapsomaniki et al., 2018). To account for the cell size, method has been developed to experimentally estimate cell size by a total protein measure. Subsequently, by normalizing measured single-cell parameters to the cell size, (relative) concentration information can be gained (Rapsomaniki et al., 2018).

## Objectives

Currently technologies have allowed systematically profiling cellular phenotypical and functional properties at single-cell resolution. Mass cytometry enables high throughput single-cell proteomic analysis with the capacity to measure more than 40 proteins or protein modifications. Using mass cytometry, intracellular signaling networks are ready to be quantified at steady and dynamic states, in health and disease. This will largely contribute to the systematic characterization of signaling responses and the understanding of signaling variances in heterogeneous populations.

Signaling transduction network is a key to many biological processes, including cell proliferation, growth, differentiation, survival, and responses to stress. Deregulated signaling networks are often cues to the initiation of abnormal cellular behaviors that can lead diseases, such as cancer. As a prime cause of signaling network aberrances, mutations on signaling protein cause protein conformational changes, altering catalytic function of the signaling protein. Independent of mutations, abnormalities in protein expression levels also change signaling network responses. However, dissimilar to mutations which lead to more discretized phenotypes, network modulation induced by protein expression changes is a continuous remodeling process as a function of protein expression level. Conventional bulk measurement techniques average measured parameters in all measured cell and are therefore incapable to analyze such protein abundance-dependent effects. Moreover, in cancer cells, the genomic instability induces high protein expression heterogeneity, leading to another layer of complexity. Thus, attempts to decipher, for example, protein overexpression-induced network states results in unprecise biological interpretations that overexpression, although has been identified with highly relevance in many cancer types, remains unclear for the signaling mechanism causing oncogenic behaviors.

The aims of this thesis are the following. First, we aim to develop a mass cytometry-based experimental method that can enable analysis of signaling protein overexpression-induced effects on signaling networks; by exploiting the gradient protein expression level resulting from transient transfection, we try to study the network modulation as a continuum, related to the expression levels of the analyzed protein. Second, as protein concentration determines cellular biochemical reaction rates, overexpressed signaling proteins can alter cellular signaling dynamics in response to extracellular stimuli, which can be quantified by our mass cytometry based approach. Third, given the capacity of mass cytometry to perform high throughput single

cells analysis, we aim to use the then established method in a kinome- and phosphatome-wide screen in which the signaling protein abundance-dependent effects on a cancer-related signaling network can be systematically quantified; novel signaling relationships and new abundance-dependent functions can be identified and assigned to the analyzed kinases and phosphatases, especially those have not been fully characterized; and overexpression-related oncogenic signaling responses can be identified.

In addition to the technical development for an protein abundance-dependency measurement approach, the thesis also aim to provide biological insights in signaling transduction and regulatory mechanisms in several major cancer-related signaling pathways, including the MAPK/ERK cascade, the AKT pathways, the PKC pathways, stress response pathways and many STAT proteins. Ultimately, our analysis can result in unique data that will facilitate computational biology analysis, including signaling network reconstruction, network modeling and mechanistic modeling.

# Result 1: Influence of node abundance on signaling network state and dynamics analyzed by mass cytometry

Xiao-Kang Lun[1,2], Vito RT Zanotelli[1,3], James D Wade[1,4], Denis Schapiro[1,3], Marco Tognetti[1,2,5], Nadine Dobberstein[1] & Bernd Bodenmiller[1]

[1] Institute of Molecular Life Sciences, University of Zürich, Zürich, Switzerland
[2] Molecular Life Science Ph.D. Program, Life Science Zürich Graduate School, ETH Zürich and University of Zürich, Zürich, Switzerland
[3] Systems Biology Ph.D. Program, Life Science Zürich Graduate School, ETH Zürich and University of Zürich, Zürich, Switzerland
[4] Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, United States of America
[5] Institute of Biochemistry, ETH Zürich, Zürich, Switzerland

Correspondence should be addressed to
B.B. (bernd.bodenmiller@imls.uzh.ch)

## Abstract

**Signaling networks are key regulators of cellular function. Although the concentrations of signaling proteins are perturbed in disease states, such as cancer, and are modulated by drug therapies, our understanding of how such changes shape the properties of signaling networks is limited. Here we couple mass cytometry-based single-cell analysis with overexpression of tagged signaling proteins to study the dependence of signaling relationships and dynamics on protein node abundance. Focusing on the epidermal growth factor receptor (EGFR) signaling network in HEK293T cells, we analyze 20 signaling proteins during a one hour EGF stimulation time course using a panel of 35 antibodies. Data analysis with BP-$R^2$, a measure that quantifies complex signaling relationships, reveals abundance-dependent network states and identifies novel signaling relationships. Further, we show that upstream signaling proteins have abundance-dependent effects on downstream signaling dynamics. Our approach elucidates the influence of node abundance on signal transduction networks and will further our understanding of signaling in health and disease.**

## Introduction

Signaling networks are at the core of cellular information processing and transform external signals into cellular responses. Signals are transduced by modulating enzymatic activities mainly via protein phosphorylation, and cells implement sophisticated mechanisms, such as feedback loops, pathway crosstalk, and differential enzyme localization, to integrate signals and drive cellular processes and physiological outputs. The abundance of individual signaling pathway components (nodes) is central to the activity and output of a signaling network (Wolf-Yadlin et al., 2006). Changes in node abundance are tightly regulated and control biological programs such as stem cell differentiation and embryogenesis (De Los Angeles et al., 2015). Abundance deregulation of particular signaling network nodes via genomic, transcriptional, or post-transcriptional regulatory defects (Bywater et al., 2013; Feinberg, 2007; Silvera et al., 2010) underlies human diseases, the prime example being cancer (Santarius et al., 2010). Copy number alterations of genes encoding critical proteins (Eralp et al., 2008; Govindarajan et al., 2007; Han et al., 2015), independent of mutations that constitutively change enzymatic activity (Davies et al., 2002), drive progression of many cancer types. Genomic instability in cancer cells causes abnormally broad distributions of signaling protein abundances in a given tumor (Wang

et al., 2015), yet the consequences of the protein abundance levels on signaling properties is poorly understood limiting our ability to rationally design therapies.

The epidermal growth factor receptor (EGFR) signaling network is affected by gene copy number alterations that deregulate protein abundances (e.g., of EGFR, HER2, ERK and AKT) in a number of cancer types (Eralp et al., 2008; Govindarajan et al., 2007; Han et al., 2015). EGFR signaling controls cell growth, motility, survival, differentiation, and metabolism (Citri and Yarden, 2006). Many drugs target the activity of the EGFR signaling network (Roberts and Der, 2007; Tebbutt et al., 2013). The receptor tyrosine kinase (RTK) function of EGFR is activated by its dimerization upon ligand binding. EGFR auto-phosphorylation recruits adaptor proteins that typically activate the MAPK/ERK and AKT signaling pathways. The MAPK/ERK branch activates the GTPase RAS, which triggers a kinase phosphorylation cascade consisting of RAF, MEK, ERK, and p90RSK. The output of the MAPK/ERK branch is transcription of genes regulating growth and division (Mendoza et al., 2011; Olayioye et al., 2000). Signal transduction through the AKT branch starts by PI3K activation, producing PIP3, which recruits AKT and PDK1 to the plasma membrane. PDK1 phosphorylates AKT (Manning and Cantley, 2007; Mendoza et al., 2011), which mediates signaling through the mTORC1 complex to modulate translation via p70S6K and 4EBP1 (Manning and Cantley, 2007). Other AKT targets are GSK3β, PRAS40, and TSC2. The AKT pathway controls cell survival, proliferation, and migration (Manning and Cantley, 2007). STAT proteins and the PKC pathway can also be activated by EGFR-mediated signaling (Bowman et al., 2000; Oliva et al., 2005). EGFR signaling involves crosstalk and feedback loops both internally (e.g., active ERK attenuates upstream RAF or MEK signaling via negative feedback) (Mendoza et al., 2011) and with other signaling pathways (e.g., WNT and TGF-β pathways) (Kim et al., 2007; Massague, 2003).

Classically, two approaches are used to characterize the effect of proteins on signal transduction. The first approach analyzes cell populations. Here, western blotting, mass spectrometry, RNA-microarrays, and synthetic lethality screens are used to identify signaling relationships (Corcoran et al., 2013; Kim et al., 2012; Zhang et al., 2005). Protein-protein interaction analyses are used to determine which proteins in a network directly interact (Kim et al., 2012; Tewari et al., 2004). Population-based methods yield a comprehensive view of signaling but are difficult to use in analysis of protein abundance dependencies due to inherent limitations: Proteins must be expressed at different abundances or cells must be sorted to yield a non-continuous abundance titration. Such methods result in a large number of samples and

cell-to-cell protein abundance variations within each sample remain masked. The second approach studies signaling relationships in single cells. Here fluorescence microscopy and flow cytometry (FACS) are used with a variety of assays, including proximity ligation assay (PLA) (Sundqvist et al., 2013) or fluorescence resonance energy transfer (FRET) (Aoki et al., 2013). These approaches allow study of signaling relationships and dynamics through time and space; however, only a few signaling nodes can be measured simultaneously.

A recently developed single-cell analysis technology, called mass cytometry, allows for the simultaneous measurement of over 40 signaling nodes in single cells using metal-isotope tagged antibodies (Bendall et al., 2011; Bodenmiller et al., 2012). This capability makes mass cytometry uniquely suited to comprehensively query the function of nodes in signaling networks within heterogeneous cell populations. Mass cytometry is quantitative and, in combination with mass-tag cellular barcoding (MCB), a powerful screening tool (Bodenmiller et al., 2012). Algorithms to analyze multiplexed single-cell mass cytometry data allow quantification of signaling relationships, therefore helping to decipher the highly complex network behaviors that operate even in simple biological systems (Krishnaswamy et al., 2014).

Here, we coupled protein overexpression with mass cytometry to measure the effect of varying node abundance on the activation state and signaling relationships of an unstimulated EGFR signaling network, as well as the signaling dynamics of the network in response to EGF stimulation. We exploited the finding that transient protein overexpression in a cell population typically produces a continuous abundance range of the target protein over four orders of magnitude. We overexpressed 20 central EGFR signaling network proteins individually in human embryonic kidney (HEK) 293T cells, sampled during an EGF stimulation time course over 60 minutes totaling 360 conditions. An average of 11,000 cells per condition was analyzed with a panel of 35 antibodies to provide a comprehensive single-cell proteomic EGFR network analysis. To identify signaling relationships in this dataset, we developed a statistical measure that we call 'binned pseudo R-squared' (BP-$R^2$) that recapitulated known signaling relationships and identified relationships that were –to the best of our knowledge- not described previously. Thus, our experimental and computational approach enables study of how the strength and dynamics of signal transduction are tuned by node abundances.

## Results

### ***Analyzing continuous protein abundance dependencies***

To systematically identify and characterize protein abundance-dependent signaling relationships, dynamics, and network activation states, we exploited the variation and large dynamic range of protein abundance induced by transient transfection and used mass cytometry to quantify the abundance of the transfected protein of interest (POI) in conjunction with comprehensive signaling network readouts in single cells. We cloned POIs genes into vectors containing a cytomegalovirus (CMV) promoter and a GFP-tag sequence (Couzens et al., 2013) to transiently overexpress GFP-tagged POIs in HEK293T cells (Fig. 1a). The tagged protein abundance was measured by mass cytometry using an anti-GFP antibody (Fig. 1a). Ordering the measured cells based on the GFP signal provided a continuous POI titration (Fig. 1b). Typically, not all cells were transfected, yielding an internal control for every experiment. To measure the single-cell EGFR signaling network states, we designed and validated a panel of 35 antibodies that mostly detect phosphorylation sites on signaling proteins (Supplementary Tables 1-3). These data were used to determine the abundance dependencies of network activation state and signaling dynamics (Fig. 1b).

To validate our system we confirmed that, first, the GFP tag was reliably detected by mass cytometry (Supplementary Fig. 1); second, the GFP tag did not affect the localization and activity of the POI (Supplementary Fig. 2, 3, Supplementary Table 4, Supplementary File 1); third, POI expression levels were linearly related to GFP abundance, validating GFP as readout of the total POI abundance (Supplementary Fig. 4a, c); fourth, POI overexpression for 18 hours (i.e., the time point of our experiments) did not alter the underlying network structure (Supplementary Fig. 4b, c); fifth, the antibody-based GFP quantification by mass cytometry was comparable to FACS (Supplementary Fig. 5); sixth, the cell culture media and cell detachment did not alter signaling processing in the EGFR network (Supplementary Fig. 6, 7); and, seventh, the levels of the GFP-tagged POIs were stable during the 1-hour EGF stimulation time course (Supplementary Fig. 8, Supplementary Video 1). We also found that the method is robust and highly reproducible as evidenced by the high concordance between the three individual experiment replicates (Supplementary Fig. 9, Supplementary File 2).

***KRAS$^{G12V}$ and MEK1$^{DD}$ abundance effect on signaling***

We first studied a well-known signaling circuit: Constitutively active mutants of KRAS and MEK1 (KRAS$^{G12V}$ and MEK1$^{DD}$) lead to ERK phosphorylation and activate components downstream in the MAPK/ERK pathway. As expected, we found that overexpression of KRAS$^{G12V}$-GFP or MEK1$^{DD}$-GFP increased phosphorylation on Thr202 and Tyr204 of ERK1/2 (Fig. 2a). Our approach also elucidated the abundance-dependent effects on these signaling relationships: The relationship between KRAS$^{G12V}$-GFP and p-ERK1/2 was bow-like as high levels of KRAS$^{G12V}$-GFP corresponded to reduced phosphorylation of ERK1/2. By contrast, the MEK1$^{DD}$-GFP abundance relationship with p-ERK1/2 was monotonic as p-ERK1/2 increased with MEK1$^{DD}$-GFP expression (Fig. 2a). These results verified the oncogenic activation of p-ERK1/2 induced by KRAS$^{G12V}$ and MEK1$^{DD}$.

Next, we analyzed the impact of KRAS$^{G12V}$-GFP and MEK1$^{DD}$-GFP abundance on all measured phosphorylation sites. We divided the measured cells into 10 bins according to the GFP signals and plotted the bin medians (Fig. 2b, Supplementary Fig. 9b-e). This analysis revealed that the phosphorylation site abundances on ERK1/2 and its direct downstream target Ser380 of p90RSK had similar relationships to the abundances of KRAS$^{G12V}$-GFP or MEK1$^{DD}$-GFP. Phosphorylation of AKT on Ser473 and its direct target Ser9 of GSK3 had parallel trends and showed reduced levels when the MAPK/ERK signal peaked, suggesting inter-pathway regulation. We also observed increased JNK phosphorylation on Thr183/Tyr185 induced by the KRAS$^{G12V}$ mutant (Fig. 2b) as reported previously (Zhou et al., 2010). This shows that our approach recapitulates known signaling relationships and identifies abundance-determined signaling responses.

We then systematically evaluated signaling relationships between all pairs of measured markers modulated by KRAS$^{G12V}$-GFP or MEK1$^{DD}$-GFP overexpression. We exploited the fact that overexpression of one protein increases signaling (i.e., phosphorylation levels) and thus expands the dynamic range of many measured markers (Fig. 2c). This enabled the use of correlation analysis to distinguish signaling relationships (high correlation) from biological and technical noise (low correlation). For example, overexpression of KRAS$^{G12V}$-GFP resulted in an increased Spearman correlation between p-ERK1/2 and p-p90RSK compared to control (Fig. 2c), whereas ERK-independent phosphorylation sites, such as Tyr551 of BTK/ITK, showed low correlation with p-ERK1/2 levels in both control and overexpression conditions (Fig. 2d).

Identifying changes in pairwise Spearman correlations for all measured markers in the $KRAS^{G12V}$-GFP and $MEK1^{DD}$-GFP overexpression data compared to the FLAG-GFP control enabled systematic analysis of signaling relationship patterns (Fig. 2e, f). Phosphorylation levels of proteins in the MAPK/ERK pathways showed strong increases in correlation, and pathway members clustered together (Fig. 2e, f, green squares). We also observed that phosphorylations of MAPK/p38 pathway members and STAT proteins (STAT1 and STAT5) were increasingly correlated with levels of MAPK/ERK pathway members as $MEK^{DD}$-GFP levels increased (Fig. 2f, purple rectangle), indicating crosstalk between MAPK and STAT pathways. These results reveal relationships among many measured markers and show that increases in correlation reflect pathways and grouped biological processes.

***Automated analysis of abundance-induced signaling***

Spearman correlation analysis can uncover strictly monotonic relationships between phosphorylation levels on signaling proteins; however, protein abundance-dependent signaling responses can be complex (Fig. 2a, see $KRAS^{G12V}$). We therefore developed a density-independent measure termed 'binned pseudo R-squared' (BP-$R^2$) to quantify the strengths of relationships between the abundance of a POI and measured phosphorylation sites. BP-$R^2$ creates 10 bins across the POI-GFP expression range and calculates the relationship strength considering bin medians and the global mean (Supplementary Fig. 10a, b, Methods, Supplementary Software). Using the BP-$R^2$ values for all negative controls, a cutoff for strong signaling relationships was determined (Supplementary Fig. 10c). Benchmarking BP-$R^2$ in identifying strong signaling relationships from the overexpression datasets showed that BP-$R^2$ outperformed methods often used for this task (Krishnaswamy et al., 2014; Redell et al., 2013) (Supplementary Fig. 11a, b). The strong relationships identified by BP-$R^2$ were plotted in a two-dimensional layout guided by canonical pathways (Fig. 2g, h). The directionality of measured signaling relationships was determined by Spearman correlation of the bin medians (Supplementary Fig. 10b, Methods). A positive correlation indicates that cells show generally increasing marker levels and a negative correlation indicates generally decreasing marker levels as POI-GFP levels increase.

Analysis of $KRAS^{G12V}$-GFP and $MEK1^{DD}$-GFP overexpression versus all measured markers using BP-$R^2$ revealed strong, positively correlated relationships of $MEK^{DD}$-GFP to downstream MAPK/ERK pathway nodes. $KRAS^{G12V}$-GFP levels, although also positively correlated with

MAPK/ERK nodes, exhibited the same, but weaker relationships (Fig. 2a, b, g, h). Together, these results suggest that feedback regulation of upstream MAPK nodes differs between the studied mutants. Additionally, this network view revealed that $MEK1^{DD}$-GFP abundance had a strong positive impact on nodes in the MAPK/p38 pathway; the previously observed $KRAS^{G12V}$-induced phosphorylation of JNK (Zhou et al., 2010) was dependent on $KRAS^{G12V}$ abundance (Fig. 2g, h). These results show that overexpression of signaling proteins, in conjunction with BP-$R^2$ and correlation analysis, identifies known relationships and is a valid platform for discovery of signaling relationships in a comprehensive and abundance-dependent manner.

***Node abundance dependency analyses of the EGFR network***

To study the node abundance dependency of signaling relationships and dynamics in the EGFR signaling network, we overexpressed 20 EGFR-related signaling proteins individually in HEK293T cells (Table 1). Each of the 20 GFP-tagged POIs was validated in previous studies (Supplementary Table 5) and in our system (Supplementary Fig. 2, 3, Supplementary File 1). 18 hours after transfection, we treated cells with EGF and quantified signaling by mass cytometry over a 60-min time course. To exclude signaling relationships caused by channel-to-channel spillover, we applied a stringent experimental filter (Supplementary Fig. 12, Methods). The median marker intensities during the time course are shown in Supplementary Fig. 13a. Based on these data we performed two sets of analyses. In the first, we used BP-$R^2$ analysis and Spearman correlations to evaluate how the abundance of overexpressed proteins influenced phosphorylation at the measured sites (Fig. 3, Supplementary Fig. 13b and Supplementary Files 2-4). In the second, we examined how features of signaling dynamics depend on protein abundance (Fig. 4).

In the first analysis, strong and broad signaling responses to overexpression were identified for the upstream kinases PDK1-, GSK3β-, SRC-, and ASK1-GFP without EGF stimulation (Fig. 3. Overall, we identified 59 strong signaling relationships in the unstimulated conditions. Overexpression of many kinases induced strong and positively correlated signaling relationships with their own phosphorylation (Fig. 3, Supplementary File 4). Overexpression of CRAF-, KRAS-, p70S6K-GFP, and others only induced signaling responses upon EGF stimulation (Fig. 3). Notably, under stimulated conditions, KRAS-, CRAF-, and MEK1-GFP levels negatively correlated with phosphorylation levels of downstream kinases p-ERK1/2 and p-p90RSK (Fig. 3). Activating mutations in KRAS and CRAF (Fig. 2), but not protein overexpression alone, may activate oncogenic signaling.

To systematically assess signaling relationships identified by BP-$R^2$, we used the literature-based signaling network, SIGNOR (Perfetto et al., 2016). For each relationship, we computed the shortest signed directed path length according to the SIGNOR network (Supplementary Table 6). We found that 76% of the strong relationships identified in the unstimulated conditions had paths with a maximum of three steps, highlighting that our approach identifies rather direct signaling relationships. Only 14 abundance-dependent relationships with four or more path steps were identified. Comparison of our strong signaling relationships with literature indicated that many EGF signaling connections that we identified were previously reported. We also propose many relationships that have—to our knowledge—not been previously reported, for example: p90RSK to PDK1 (Ser241), GSK3β to SHP2 (Tyr580), JNK1 to MAPKAPK2 (Thr334), p110α to MKK3 (Ser189), p110α to MKK6 (Ser207), ASK1 to PDK1 (Ser241), ASK1 to GSK3β (Ser9), and ASK1 to AMPKα (Thr172) (Table 2).

Phosphorylation levels of many members of the MAPK/ERK pathways showed complex relationships (i.e., measured phosphorylation levels varied over the analyzed POI-GFP range and the relationships did not fit linear, sigmoidal, or quadratic models) with levels of POI-GFPs upon EGF stimulation. These relationships can be explained by abundance-dependent modulation of the signaling dynamics in response to EGF. Thus, in the second set of analyses we examined how signaling dynamics, as quantified by amplitude and peak-time, depended on abundance of an overexpressed protein (Fig. 4). In order to view signaling trajectories as functions of protein abundance, we binned the POI-GFP levels into 10 bins (Fig. 4a, Supplementary File 2). This allowed tracing the signaling trajectories of cells with similar protein overexpression levels (i.e., those in the same bin) over the EGF stimulation time course (Fig. 4b, Supplementary File 5). Strong and robust changes in signaling amplitudes (Fig. 4c-i) and peak-times (Supplementary Fig. 14) were found. Notably, the maximum amplitudes were independent of the overexpression range of a given POI (Supplementary Fig. 15).

We found that high CRAF-GFP and KRAS-GFP abundance strongly reduced signaling amplitudes of p-ERK1/2 and p-p90RSK (Fig. 4c, d, i), whereas high abundance of MEK1-GFP strongly reduced amplitudes and delayed peak-times for p-p90RSK (Fig. 4i, Supplementary Fig. 14). Overexpression of ERK2-GFP led to complex abundance-dependent responses of p-p90RSK and p-ERK1/2 after EGF stimulation (Fig. 4e-h). p-ERK1/2 amplitudes increased and peak-times delayed as a function of ERK2-GFP abundance level (Fig. 4g-i, Supplementary Fig.

14). Intermediate abundance levels of ERK2-GFP also delayed the p-p90RSK peak-times relative to low ERK2-GFP abundance, whereas cells with high ERK2-GFP levels exhibited minimal p-p90RSK signaling dynamics (Fig. 4e, f, i, Supplementary Fig. 14). Overexpression of p90RSK-GFP modulated the signaling amplitude of its potential crosstalk phosphorylation site, Ser241 of PDK1, in an abundance-dependent manner, and increasing expression of p90RSK increased p-PDK1 amplitudes (Fig. 4i). Thus, we observed abundance-dependent signaling dynamics across the range of overexpression levels. Overexpression of upstream signaling proteins (KRAS-, CRAF-, MEK1-, and ERK2-GFP) in the MAPK/ERK pathway led to reduced signaling amplitudes and delayed peak-times of their downstream targets. These observations show that our approach can quantify the role of protein abundance in determining the dynamic signaling response to an extracellular stimulation.

## Discussion

Here we present an approach coupling transient overexpression with mass cytometry-based single-cell measurements to characterize signaling network activation states and signaling dynamics over a quasi-continuous, high dynamic range of protein abundance. To highlight the utility of our approach, we present a comprehensive single-cell proteomic analysis of the EGFR network that enabled an analysis of abundance-dependent effects of signaling proteins on state and dynamics of the signaling network. We evaluated the effects of overexpressing 20 EGFR network key nodes with a 60-minute EGF stimulation time course. In each of the 360 conditions, we measured the effect of a POI over a four order of magnitude abundance range on 35 markers by mass cytometry providing a unique and valuable quantitative single-cell resource of abundance dependencies of EGFR signaling.

Previously, the heterogeneity of protein levels after transient transfection was considered problematic. Here, we took advantage of this cell-to-cell variation as it results in a continuous titration of protein abundance over four orders of magnitude. Untransfected cells also provided an internal control for each experiment. We used the multiplexing capabilities of mass cytometry to characterize abundance dependencies of signaling network state and dynamics. Applied to the EGFR signaling network, our approach recapitulated known relationships, suggested previously not described ones, and revealed the intricate modulation of signal amplitudes and peak-times as functions of continuous protein abundance.

Our approach contributes to the understanding of signaling on several levels. First, the approach can be used to study uncharacterized proteins and to suggest additional roles to characterized ones. Second, we were able to directly relate POI abundance with the comprehensive analysis of signaling dynamics in response to stimulation. Such analyses are necessary for understanding of differential signal processing in identical cell types and in disease states characterized by heterogeneity in protein expression such as cancer. Third, the overexpression yields a large dynamic range of signaling activity and can reveal signaling relationships masked by stochastic processes and technical noise under otherwise similar conditions, facilitating the computational analysis of signaling relationships. Fourth, we present a metric termed BP-$R^2$, which allows the quantification of the strengths of arbitrary shaped signaling relationships. BP-$R^2$ was superior to state-of-the-art methods for analysis of our dataset. Fifth, and finally, we were able to infer protein abundance-dependent signaling kinetics from single-cell snapshot data.

Our approach recapitulated known oncogenic signaling behaviors induced by the constitutively active mutants $KRAS^{G12V}$ and $MEK1^{DD}$ and identified novel abundance-dependent signaling relationships. For example, p-ERK1/2 was attenuated in cells with highly overexpressed $KRAS^{G12V}$-GFP, potentially due to negative feedback loops or senescence (Xu et al., 2014). Overexpression of the wild-type KRAS-GFP and MEK1-GFP did not induce downstream signaling activation, suggesting that mutations on KRAS or MEK1 are the main drivers of oncogenic signaling. Further, our approach allows study of abundance-dependent signaling dynamics. In the MAPK/ERK pathway, high abundance of upstream signaling mediators KRAS, CRAF, MEK1, or ERK2 reduced amplitudes and delayed peak-times of downstream phosphorylation sites. One possible explanation is that the signal transduction is determined by the competition between active and inactive forms of a signaling protein for substrates. Overexpression increases the total abundance but may reduce the percentage of the active form.

*KRAS* amplification has been identified in many cancer types. Amplification, however, is not correlated with the phosphorylation of ERK1/2 (Rahman et al., 2013). Rather, *KRAS* amplification mediates resistance to inhibitors targeting growth pathway related kinases, including EGFR, MET and MEK1/2; KRAS knockdown diminishes the drug resistance (Cepero et al., 2010; Little et al., 2011; Valtorta et al., 2013). Our results indicate that due to reduced downstream signaling amplitudes in response to EGF stimulation, the dependency of cells on

the MAPK/ERK pathway may decrease upon KRAS overexpression, suggesting a mechanism for cancer cell resistance to inhibitors.

Comparing the identified strong signaling relationships with those in the SIGNOR database, we propose previously not described signaling relationships, e.g.: 1) Our data suggest that p90RSK potentially forms a positive feedback loop and activates the upstream signaling protein PDK1. 2) GSK3β has been identified as a central signaling controller and has multiple substrates (Cohen and Frame, 2001); our results suggest that SHP2 is a potential direct or indirect target of GSK3β. 3) We also propose that JNK1 is a MAPKAPK2 activator. 4) PI3K and MKK3/6 are known to be regulated by RAC1 (Shin et al., 2005); our results suggest PI3K activates MKK3/6 independently. 5) Recent studies indicate that ASK1 contributes in negative regulation of PDK1 through phosphorylation on Thr254 of PDK1 (Seong et al., 2010); We observe ASK1 overexpression-induced PDK1 phosphorylation on Ser241, inducing PDK1 activity and downstream GSK3β phosphorylation on Ser9. 6) In addition to the known AMPK-mediated ASK1 activation (Lee et al., 2010), our data indicates ASK1 activation of AMPKα via phosphorylation on Thr172. 7) We have also observed negative correlations between the abundance of p70S6K or PDK1 to the phosphorylation level of S6 (Ser235/Ser236), indicating overexpression-induced-negative feedback regulations.

Our method has several limitations. First, we do not measure the endogenous expression level of the POI. However, exogenous expression is linearly correlated with the total protein level (Supplementary Fig. 4a), validating GFP as readout of the total POI. Second, all results in mass cytometry rely on antibodies; for this work, all antibodies were thoroughly validated (Supplementary Table 3). Third, we do not measure the abundance range of the studied proteins in cancer cells, however, proteome studies of cancer cells and databases such as PaxDb (Wang et al., 2015) indicate a range similar to those studied here. Fourth, high expression levels of a protein kinase may induce non-specific phosphorylation; however, our data allows choosing the analyzed expression range *in silico*, thus such effects can be excluded.

The approach described here provides a method to study how the abundance variance of signaling proteins in different tissues and cell lines results in distinct signaling behaviors. The application of our approach to synthetic biology, stem cell biology, developmental biology, and cancer-related processes, such as the epithelial-mesenchymal transition, will enable quantitative identification of key proteins and signaling determinants in cell differentiation at

phenotypical switching points. We envision that determining which signaling relationships and thresholds enable diseased cells to overcome drug treatment will be a highly relevant application.

## Accession codes

All data and working illustrations are available on Cytobank.org under the project 725. The BP-$R^2$-based analysis is provided as Supplementary Software and the GitHub repository https://github.com/BodenmillerGroup/Adnet.


## Acknowledgements

We would like to thank the Bodenmiller lab for support and fruitful discussions, the Lehner lab and the Mosimann lab for sharing equipment. We would especially like to thank Dr. A-C. Gingras, Lunenfeld-Tanenbaum Research Institute, for sharing the pDEST vectors used in this study. This work was supported by the Swiss National Science Foundation (SNSF) R'Equip grant 316030-139220, a SNSF Assistant Professorship grant PP00P3-144874, a Swiss Cancer League grant, the PhosphonetPPM SystemsX grant, and funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 336921. The work of J.D.W. was supported by a National Science Foundation Graduate Research Fellowship under Grant No. DGE-1148903 and a Whitaker International Fellowship awarded by the Institute of International Education. The work of D. S. is supported by the Forschungskredit of the University of Zurich Fellowship.


## Author contributions

X.K.L. and B.B. conceived and designed the experiments. X.K.L., M.T. and N.D. performed experiments. V.Z., J.D.W., X.K.L. and D.S. performed data analysis. X.K.L. and B.B. wrote the manuscript. All authors commented on and edited the final version of the paper.

## Competing financial interest

The authors declare no competing financial interest.

## Figure legends

***Figure 1*:** Workflow of abundance-dependent network analysis. (**a**) Experimental workflow. Signaling POIs are cloned into vectors containing a CMV promoter and a GFP-tag sequence to transiently overexpress GFP-tagged POIs in HEK293T cells. We quantify anti-GFP antibody as readout of POI-GFP abundance, together with other 35 markers, by mass cytometry. (**b**) Data analysis workflow. Cells were ordered based on the GFP signal, providing a continuous POI titration, which was then coupled to other signaling markers to determine the abundance dependencies of network activation state and signaling dynamics in the network after transfection. The network in the illustration does not represent an actual biological example**.**

***Figure 2*:** MAPK/ERK pathway mutants induce oncogenic signaling. (**a**) Biaxial plots of GFP, representing the abundance of the overexpressed mutant POIs, versus abundance of phosphorylation on Thr202/Tyr204 on ERK1/2. Constitutively active $KRAS^{G12V}$-GFP shows a downregulation on Thr202/Tyr204 on ERK1/2 at the highest levels of $KRAS^{G12V}$-GFP. Constitutively active $MEK1^{DD}$-GFP directly phosphorylates Thr202/Tyr204 on ERK1/2, and the abundance of the POI-GFP is correlated with amount of ERK1/2 phosphorylated at these sites. The FLAG-GFP control does not affect ERK phosphorylation sites. (**b**) The abundances of measured phosphorylation sites are plotted over the range of the $KRAS^{G12V}$-GFP and $MEK1^{DD}$-GFP expression. Phosphorylation sites of the same pathway (e.g., on ERK1/2 and p90RSK, AKT and GSK3β, or p38 and JNK) show similar trends. An individual experiment is shown here. Plots for 3 replicates are shown in Supplementary Fig. 9b-e. (**c**) Strong single-cell correlations within biaxial plots indicate co-regulated phosphorylation sites. (**d**) Unchanged and reduced correlations indicate unrelated phosphorylation sites. (**e**) and (**f**) Heat maps showing for all pairs of measured markers the change in Fisher-transformed Spearman correlation values for overexpression of (**e**) $KRAS^{G12V}$-GFP and (**f**) $MEK1^{DD}$-GFP when compared to the FLAG-GFP overexpression control. (**g**) and (**h**) BP-$R^2$ scores and Spearman correlations of bin medians for all measured markers in cells where (**g**) $KRAS^{G12V}$-GFP or (**h**) $MEK1^{DD}$-GFP was overexpressed overlaid on a literature-based graph of canonical signaling pathways (Cardaci et al., 2012; Hendriks et al., 2014; Kim et al., 2012; Massague, 2003; Mendoza et al., 2011; Mitra et al., 2005; Nyati et al., 2006; Rawlings et al., 2004; Roberts and Der, 2007; Xu et al., 2014). Strong relationships identified from the BP-$R^2$ analysis are plotted on the signaling maps as colored circles. The sizes of circles indicate relationship strengths quantified by BP-$R^2$. The

directionalities of relationships, as judged by Spearman correlation of bin medians, are shown by the color of the circles (positive correlation indicates that cells show generally increasing marker levels, and a negative correlation indicates decreasing marker levels as POI-GFP levels increase). For (e) to (h), data from 3 individual experiment replicates were used.

***Figure 3*:** Analysis of dynamics of EGFR signaling. HEK293T cells overexpressing GFP-tagged signaling proteins listed in Table 1 were treated with EGF for 0, 5, 15, 30, and 60 min. Strong abundance-dependent signaling relationships (Supplementary Fig. 10c) are plotted on the signaling map with circle sizes and colors indicating strengths (BP-$R^2$ score) and directionalities (Spearman correlation of bin medians), respectively. The miniaturized network is the same as used in Fig. 2. Overexpression of S6-GFP did not induce any strong signaling relationships (data not shown). For all analyses, data from 3 individual experiment replicates were used.

***Figure 4*:** Analysis of node abundance-dependent EGFR signaling dynamics. (**a, b**) Schematic plots of amplitude and peak-time analysis. (**a**) The x-axis (i.e., overexpressed protein as determined by the GFP measurement) was split into 10 bins. (**b**) Median phosphorylation abundance in each bin was plotted on the y-axis versus time (x-axis) to visualize abundance dependency of signaling dynamics. (**c, d**) Mass cytometry ion counts (arcsinh transformed, Methods) measured for p-p90RSK (y-axis) as a function of ion counts measured for abundance of CRAF-GFP (x-axis) and EGF stimulation time. The same layouts for (**e, f**) ERK2-GFP abundance-determined p-p90RSK levels and (**g, h**) p-ERK1/2 levels are shown. (**i**) Heat map showing protein abundances with strong influences on signaling amplitudes with color indicating normalized signaling amplitudes. Only overexpressed proteins with an amplitude-ratio higher than 3 fold for more than two of the three replicates were identified as strong influences and are included in the heat map. For (a) to (h), representative examples from the 3 individual experiment replicates are shown. Other replicates are presented in Supplementary File 5. In (i), all replicate data are shown.

**Table 1: Overexpressed signaling proteins**

| Overexpressed proteins | Gene ID | UniProt Entry |
|---|---|---|
| SRC | *SRC* | P12931 |
| PDK1 | *PDPK1* | O15530 |
| AKT1 | *AKT1* | P31749 |
| GSK3β | *GSK3B* | P49841 |
| MKK7 | *MAP2K7* | O14733 |
| MKK6 | *MAP2K6* | P52564 |
| p38α | *MAPK14* | Q16539 |
| ERK2 | *MAPK1* | P28482 |
| p90RSK | *RPS6KA1* | Q15418 |
| CRAF | *RAF1* | P04049 |
| JNK1 | *MAPK8* | P45983 |
| p110α | *PIK3CA* | P42336 |
| BRAF | *BRAF* | P15056 |
| ASK1 | *MAP3K5* | Q99683 |
| p70S6K | *RPS6KB1* | P23443 |

| | | |
|---|---|---|
| MEK1 | *MAP2K1* | Q02750 |
| KRAS | *KRAS* | P01116 |
| HRAS | *HRAS* | P01112 |
| SHP2 | *PTPN11* | Q06124 |
| S6 | *RPS6* | P62753 |

**Table 2: Relationships with shortest singed directed path length above 3 in the SIGNOR database**

| Overexpressed POI | Target | Sign | Shortest Signed Directed Path (SIGNOR) | Literature Information |
|---|---|---|---|---|
| SRC | p-BTK/ITK | 1 | 6 | SRC family kinases phosphorylate BTK (Hendriks et al., 2014) |
| SHP2 | p-S6 | -1 | 5 | Known regulation (Marin et al., 2008) |
| ASK1 | p-PDK1 | 1 | 5 | Potential novel relationship |
| SRC | p-PLCγ2 | 1 | 5 | SRC family kinases activates PLCγ2 (Hendriks et al., 2014) |
| ASK1 | p-AMPKα | 1 | 4 | Potential novel relationship |

| | | | | |
|---|---|---|---|---|
| GSK3β | p-SHP2 | 1 | 4 | Potential novel relationship |
| p90RSK | p-PDK1 | 1 | 4 | Potential novel relationship |
| JNK1 | p-STAT1 | 1 | 4 | JNK activates STAT1 (Wei et al., 2014) |
| JNK1 | p-MAPKAPK2 | 1 | 4 | Potential novel relationship |
| p110α | p-MKK3/6 | 1 | 4 | Potential novel relationships |
| HRAS | p-SMAD2/3 | 1 | 4 | Known crosstalk (Massague, 2003) |
| ASK1 | p-GSK3β | 1 | 4 | Potential novel relationships |
| PDK1 | p-S6 | -1 | 4 | Overexpression-induced negative regulation |
| p70S6K | p-S6 | -1 | 4 | Overexpression-induced negative regulation |

## Methods

### ***Data availability***

All raw data are available at http://www.cytobank.org/bodenmillerlab.

### ***Cloning***

DNA sequences of the genes of interest were provided in entry clones by William Hahn and David Root(Yang et al., 2011) (via Addgene and NEXUS Personalized Health Technologies at ETH Zurich). Destination vectors, including pDEST pcDNA5 FRT TO-eGFP, pDEST 5' Triple Flag pCDNA5 FRT TO and pDEST 3' Triple Flag pcDNA5 FRT TO, were kindly provided by Dr. Anne-Claude Gingras at Mount Sinai Hospital, Toronto, Canada(Couzens et al., 2013). Expression vectors encoding the FLAG- or GFP-tagged fusion proteins were generated via Gateway Cloning and sequenced before transfection.

### ***Cell culture***

HEK293T cells, obtained from ATCC, were cultured in DMEM (D5671, SIGMA), supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin. For cell passaging or harvesting, cells were incubated with 1X TrypLE™ Express (Life Technologies) for 2 minutes at 37 °C.

### ***Transfection and stimulation***

HEK293T cells were seeded at the density of 0.7 million per well in 6-well plates. After 24 hours, cells were transfected with 2 µg plasmid and 4 µl of jetPRIME (PolyPlus) per well with the standard protocol provided by the manufacturer. At 18 hours after transfection, EGF (Peprotech) was added to a final concentration of 100 ng/ml. At 20 minutes before a given EGF stimulation time point, 5-iodo-deoxycytidine (IdU) was added to the medium at the final concentration of 10 µM. At 2 minutes before a given EGF stimulation time point, medium was replaced by 1X TrypLE to induce cell detachment. At the time point, paraformaldehyde (PFA, from Electron Microscopy Sciences) was added to the cell suspension to a final percentage of 1.6%, and cells were incubated at room temperature for 10 minutes. If EGF stimulation was not necessary in the experiment, cells were directly harvested and crosslinked with PFA. Crosslinked cells were washed twice with cell staining media (CSM, PBS with 0.5% BSA, 0.02% $NaN_3$) and after centrifugation, ice-cold methanol was used to resuspend the cells, followed by a 10-minute

permeabilization on ice or for long-term storage at -80 °C. Three biological replicates were performed for each experiment. In each replicate, the experimental procedures were performed on different days.

***Cell sorting***

HEK293T cells overexpressing FLAG-GFP were detached from the plates as described above and resuspended in the FACS buffer (PBS with 0.5% BSA and 5 mM EDTA). Cells were sorted with BD FACSAria III Cell Sorter into GFP low, intermediate, and high levels with the strategy indicated in Supplementary Fig. 5.

***Live cell imaging***

HEK293T cells were seeded in CultureWell™ Chambered Coverglass (Thermo Fisher Scientific) pre-coated with fibronectin. Transfection of FLAG-GFP was performed as described above. At 18 hours after transfection, cells were imaged with a Leica DMI 6000 inverted microscope at 37 °C and 5% $CO_2$. Images were taken every 2 minutes for 1 hour.

***Immunofluorescence staining***

Cells were cultured in 16-well glass chamber slides (Thermo Fisher Scientific). Transfection was done as described above. Before staining, culture medium was removed and the slide was then washed with PBS. To crosslink cells, 4% PFA was added, and cells were incubated at room temperature for 20 minutes. Slides were washed with PBS three times, and cells were permeabilized for 5 minutes with 0.1% TritonX-100 dissolved in PBS at room temperature. After washing three times with PBS, cells were incubated in blocking buffer (10% goat serum diluted in PBS) for 30 minutes at room temperature. Primary (anti-GFP, FM264G, BioLegend, 1:200) and secondary (Goat anti-Rat Alexa Fluor® 568, 1:500, supplemented with Hoechst 33342 at a final concentration of 100 µg/ml) antibodies were diluted in blocking buffer and applied to slides. A total protein stain of Alexa Fluor 647 Succinimidyl Ester (Life Technologies) was used to indicate cell outlines. Cells were washed three times with PBS after each incubation step. Slides were mounted with ProLong Gold Antifade Reagent (Life Technologies) before imaging with a CLSM Leica TCS SP8 microscope.

***Antibody conjugation***

The MaxPAR antibody conjugation kit (Fluidigm) was used to generate isotope-labeled antibodies using the manufacturer's standard protocol. After conjugation, the antibody yield was determined based on absorbance of 280 nm. Candor PBS Antibody Stabilization solution (Candor Bioscience GmbH) was used to dilute antibodies for long-term storage at 4 °C.

***Barcoding and staining protocol***

Formalin-crosslinked and methanol-permeabilized cells were washed three times with CSM and once with PBS. Cells were incubated in PBS containing barcoding reagents ($^{102}$Pd, $^{104}$Pd, $^{105}$Pd, $^{106}$Pd, $^{108}$Pd, $^{110}$Pd, $^{113}$In and $^{115}$In) at a final concentration of 100 nM for 30 minutes at room temperature and then washed three times with CSM(Bodenmiller et al., 2012). Barcoded cells were then pooled and stained with the metal-conjugated antibody mix (Supplementary Table 1) at room temperature for 1 hour. The antibody mix was removed by washing cells three times with CSM and once with PBS. For DNA staining, iridium-containing intercalator (Fluidigm) diluted in PBS with 1.6% PFA was incubated with the cells at 4 °C overnight. On the day of the measurement, the intercalator solution was removed, and cells were washed with CSM, PBS, and dd$H_2O$. After the last washing step, cells were resuspended in dd$H_2O$ and filtered through a 70-µm strainer.

***Mass cytometry analysis***

EQ$^{TM}$ Four Element Calibration Beads (Fluidigm) were added to cell suspensions in a 1:10 ratio (v/v). Samples were analyzed on a CyTOF2 (Fluidigm). The manufacturer's standard operation procedures were used for acquisition at a cell rate of ~500 cells per second. After the acquisition, all FCS files from the same barcoded sample were concatenated(Bodenmiller et al., 2012). Data were then normalized, and bead events were removed(Finck et al., 2013) before doublet removal and de-barcoding of cells into their corresponding wells using a doublet-filtering scheme and single-cell deconvolution algorithm(Zunder et al., 2015). Subsequently, data was processed using Cytobank (http://www.cytobank.org/). Additional gating on the DNA channels ($^{191}$Ir and $^{193}$Ir) and $^{139}$La/$^{141}$Pr was used to remove remained doublets, debris and contaminating particulate.

## ***Data visualization and analysis***

### *Bi-axis scatter plots*

Bi-axis scatter plots were generated in Cytobank (http://www.cytobank.org/).

### *Data preprocessing*

Raw data was transformed using the inverse hyperbolic sine transform with a cofactor of 5:

$$data = arsinh(data_{raw}/5)$$

Except where use of raw data values is specifically noted, all visualizations and analyses were performed using transformed data.

### *Data binning*

For data binning, the range between the lower and upper 2.5% of observations was divided into ten equal bins $bin_1$,...,$bin_{10}$. The observations in the lower and upper 2.5% were assigned to the lowest and highest bins, respectively. In order to be able to compare expression levels between samples within a time course replicate, all observations of the time course were used to determine the binning.

### *Correlation analysis*

Spearman correlation ($r_{ij}$) was calculated between all marker pairs *i*, *j* for each replicate and condition. Fisher's z-transformation:

$$z_{i,j} = artanh(r_{ij})$$

was used to compare pairwise correlation coefficients across conditions. For each overexpression, the change in correlation matrix was calculated by subtracting the median $\tilde{z}_{ij}$ value (across replicates) of the FLAG-GFP controls from the median $\tilde{z}_{ij}$ value (across replicates) of the overexpression condition.

$$\Delta\tilde{z}_{ij} = \tilde{z}_{ij\ overexpression} - \tilde{z}_{ij\ FLAG-GFP}$$

The resulting matrix of differences in Fisher transformed correlation values was hierarchically clustered using the Ward method and Euclidean distances(Ward, 1963).

*BP-$R^2$*

Relationships between overexpression levels and markers can include non-monotonic relationships that are not properly captured with correlation metrics such as Spearman correlation. Furthermore, although the shapes induced by an overexpression were highly reproducible, the number of cells with a given expression intensity level were not. Thus, in order to quantify the strength of arbitrarily shaped relationships between markers and overexpression levels over the whole overexpression range, a density agnostic metric termed binned pseudo R-squared (BP-$R^2$) was developed. For this metric, the middle 95% of POI-GFP levels over a time-course replicate was divided into 10 equal-width bins. Bins with less than 25 cells were discarded. For each bin $i$, the median of a measured marker ($\tilde{y}_i$) was calculated. Additionally, the overall mean of the medians of all the 10 bins ($\mu_{\tilde{y}}$) was calculated. Then, for each bin, we computed the sum of squared deviations from the bin medians and the sum of squared deviations from the overall mean of medians These values were summed over all bins and the BP-$R^2$ was defined as one minus the ratio between them:

$$R^2_{BP} = 1 - \frac{\sum_{i=1}^{nbins} \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \tilde{y}_i)^2}{\sum_{i=1}^{nbins} \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \mu_{\tilde{y}})^2}$$

Following the rationale of classical R-squared statistics, BP-$R^2$ quantifies the average reduction in squared deviations per bin when modeling the data as piecewise constant within each bin (based on the bin medians) compared to using the mean over all bin medians. The BP-$R^2$ metric represents the relationship strength between a marker and the overexpressed signaling protein relative to the overall variability of the marker. By using the median instead of mean, the BP-$R^2$ selects unimodal relationships with low noise over noisy, multimodal relationships. Notably this measure works with arbitrary interaction shapes and is largely robust against density in homogeneities. In order to aggregate the sample replicates, we considered the median BP-$R^2$ value across the experimental triplicates.

*Threshold determination*

We observed many relationships with low BP-$R^2$ between overexpressed proteins and measured phosphorylation markers, even within control samples (overexpression of FLAG-GFP). We proposed such weak relationships are more likely to result from indirect biological mechanisms; therefore, we focused on relationships that were stronger than all relationships seen in the controls (FLAG-GFP overexpression and untransfected cells). We chose the maximum median (across replicates) BP-$R^2$ of all controls (FLAG-GFP overexpression and untransfected cells) as a cutoff. Relationships that had a median BP-$R^2$ higher than this threshold were considered as sufficiently strong to be of interest.

*Kinetic analysis*

For each overexpression condition, replicates of EGF stimulation time courses were processed, stained, and measured together. Simultaneous processing enabled direct quantitative comparisons of the measured POI-GFP counts in these time courses. Samples representing all time points in a time course replicate were combined and binned by POI-GFP intensity as described in the *data binning* section. As the binning was performed over all samples of the same time course, the range of GFP intensity of bins with the same bin index directly corresponds to cells with similar abundance levels of POI-GFP in each of the different time points. As POI-GFP levels stay quasi-constant over the timescale of the 60-minute time course (Supplementary Fig. 8, Supplementary Movie 1), tracking how the median marker levels in a specific bin change over the time course reflects the kinetics of cells with a similar abundance level upon stimulation. Thus, the kinetic responses over a range of low-abundance to high-abundance cells can be compared and analyzed using classical signal processing readouts such as signal response amplitude and peak-time. For this analysis, only POI-GFP marker pairs with at least one strong relationship over the time course were considered.

*Amplitude analysis*

The response amplitude for each binned abundance level was calculated using raw counts. For each measured marker and time point, the median marker level of each POI-GFP bin was divided by the median level of the marker in the corresponding bin of the unstimulated sample (EGF 0 min) to calculate amplitude as a fold change. The amplitude for each bin was identified as the maximum fold change over all time points. Robust and strong abundance-dependent changes were identified by comparing the amplitude-ratio between the second highest and the

second lowest bin-amplitude. Of those identified as robust and strong, overexpressed proteins with an amplitude-ratio higher than 3 folds for more than two of the three replicates were identified as interesting and plotted as a heat map.

*Peak-time analysis*

Interesting examples for overexpression changes in bin peak-time were defined by identifying the time point with maximum amplitude for each bin as the peak-time. Consistent and robust examples were selected by the following criteria: Monotonically increasing or decreasing peak-times over the increasing overexpression bins and a clear change in amplitude (>3 fold) in at least 2 of the 3 replicates.

*Signor database comparison*

To compare the consistency between the strong signaling relationships detected by BP-$R^2$ analysis with the relationships predicted by the Signor database(Perfetto et al., 2016), the python NetworkX(Hagberg et al., 2008) package was used to construct a signed directed Signor network from the UniProt entries of overexpressed POIs and measured phosphoproteins. NetworkX also calculates a shortest path length within the signed directed network. Antibodies may bind to the same phosphorylation sites on more than one protein that belong to a family, making the mapping between antibodies and UniProt entries ambiguous. In this case, the shortest path value was calculated between the overexpressed POI and any possible antibody targets.

The analysis was preformed including the directionalities of signaling relationships as identified by the Spearman correlation of the bin medians in our analysis, and by exploiting the SIGNOR annotations in the following way: Simple paths between the overexpressed protein and the targeted phosphorylation sites were analyzed, starting from the shortest path to longer paths, until a sign-consistent path was found. To identify sign consistency, all edges in the SIGNOR network were classified as positive, negative or ambiguous based on the SIGNOR 'Effect' annotation: down-regulates...=negative, up-regulates...=positive, something else=ambiguous. In case there were multiple interaction types possible for an edge (positive and negative), the overall sign was taken to be ambiguous. In cases where the last edge was annotated to be affecting exactly the residue (SIGNOR annotation 'Residue') measured by the phospho specific

antibody through (de)phosphorylation (SIGNOR annotation 'Mechanism'), the directionality sign of this edge was determined to be phosphorylation=positive, dephosphorylation=negative or the inverse in cases the antibody was measuring the non-phospho site (e.g., Ser33/37/Thr41 on β-catenin). Measured phosphorylation sites responsible for inactivating a protein (e.g., Ser9 on GSK3β) were also signed as phosphorylation=negative.

A path was determined sign consistent if the product of the signs of all its edges were in accordance with the relationship direction as measured by the spearman correlation over the bins or ambiguous.

***Systematic spillover exclusion***

A stringent spillover filter was applied to systematically remove strong signaling relationships potentially affected by channel-to-channel spillover: For any measured channel that had events with ion counts over 500, we checked for spillover due to: first, isotope impurity (channels with isotopes of the same metal); second, mass resolution (-1 and +1 channels); third, oxidation (+16 channels). Any strong relationships (BP-$R^2$) with GFP and markers from these sets of channels were selected for additional verification experiments, in which the staining was done in 3 groups:

1. All antibodies in a set
2. All antibodies in a set except for the one that potentially causes spillover
3. Only the antibody potentially causing spillover

When spillover-induced background contributed to over 10% of the actual ion counts, the channel was discarded from further analysis (Supplementary Fig. 13).

Based on our spillover exclusion protocol, we found the following channels were affected by spillover. They were excluded from the analysis performed in this manuscript, and should not be considered in any subsequent analyses using this data:

In SRC overexpression:

142Nd p-SHP2

159Tb p-SMAD1/5

In PDK1 overexpression:

142Nd p-SHP2

143Nd p-FAK

145Nd p-MAPKAPK2

160Gd p-MKK3/MKK6

162Dy p-BTK/ITK

In GSK3β overexpression:

146Nd p-p70S6K

In p90RSK overexpression:

142Nd p-SHP2

147Sm p-MKK3

162Dy p-BTK/ITK

164Dy p-SMAD2/3

## Figures

Figure 1

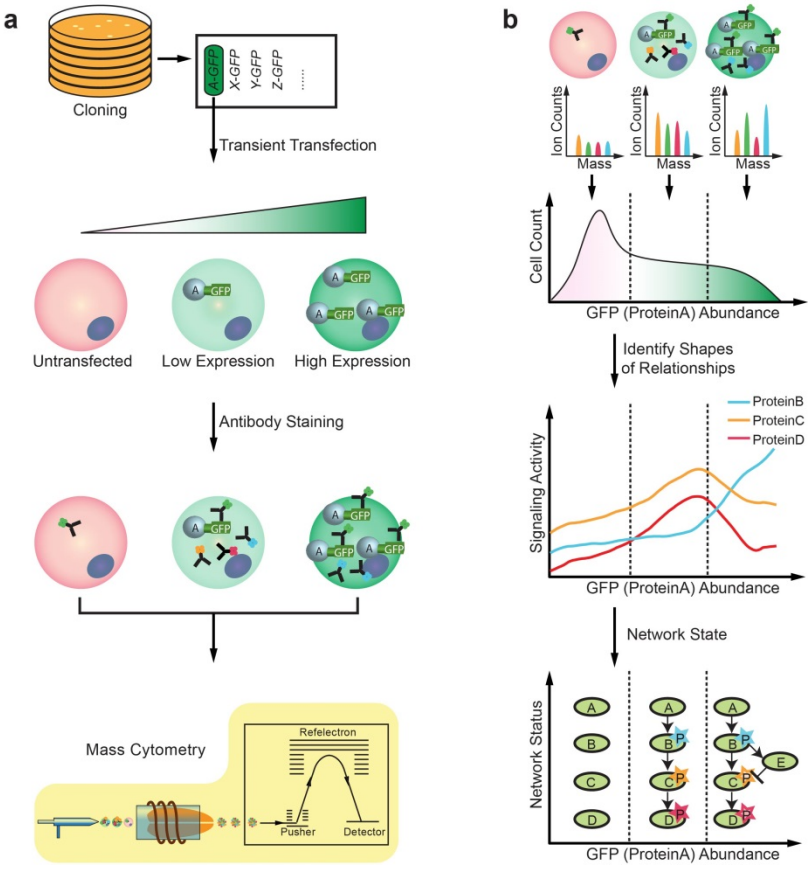


Figure 2

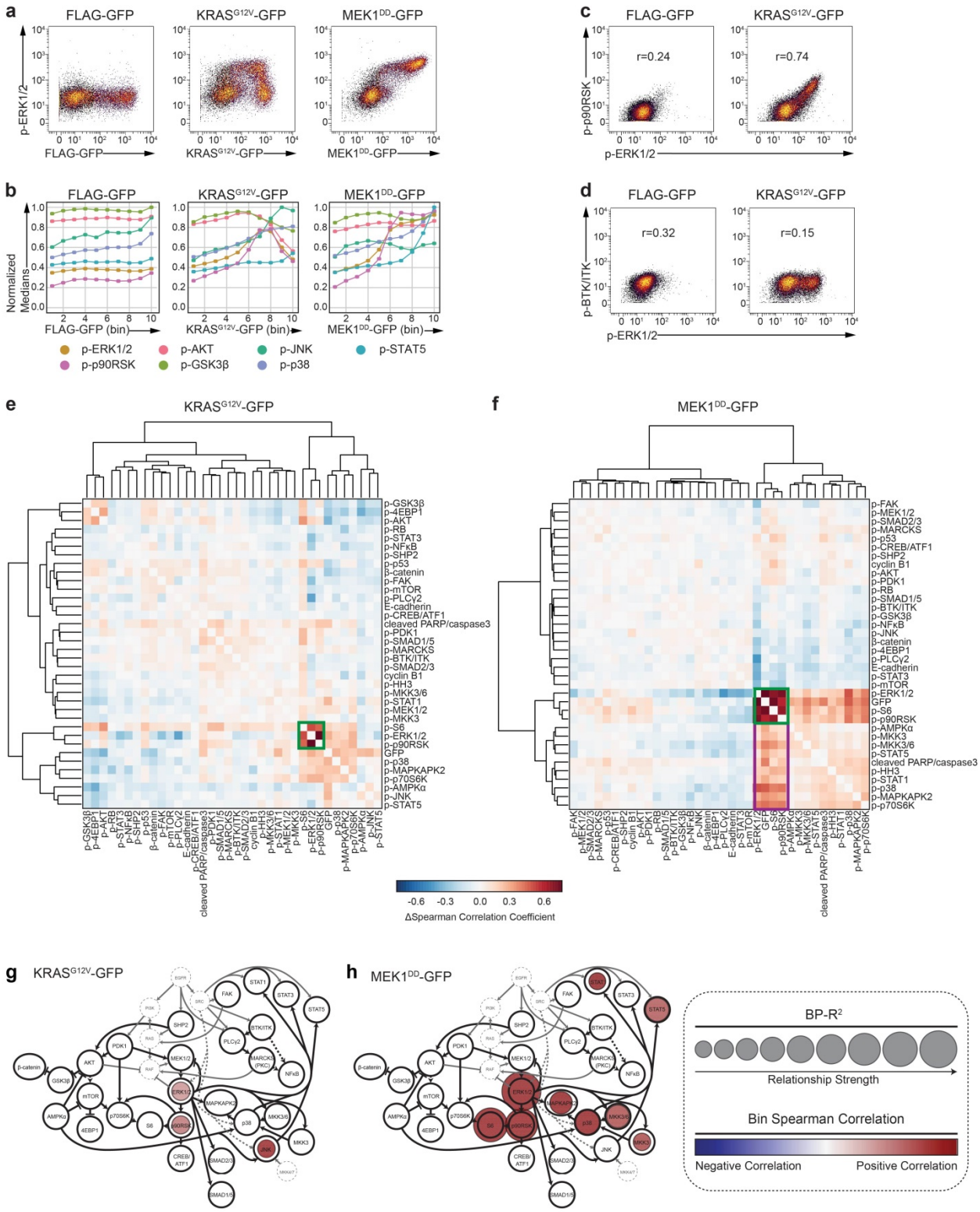
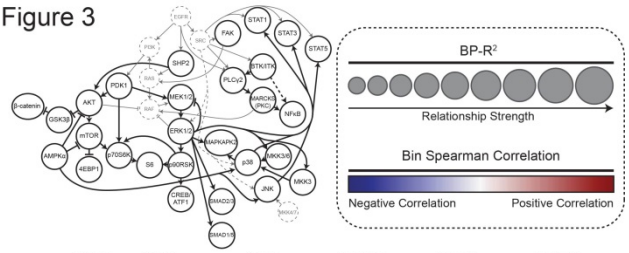

a
FLAG-GFP
KRASG12V-GFP
MEK1DD-GFP
p-ERK1/2
b
Normalized Medians
FLAG-GFP (bin)
KRASG12V-GFP (bin)
MEK1DD-GFP (bin)
p-ERK1/2
p-p90RSK
p-AKT
p-GSK3β
p-JNK
p-p38
p-STAT5
c
r=0.24
r=0.74
p-p90RSK
p-ERK1/2
d
r=0.32
r=0.15
p-BTK/ITK
e
f
ΔSpearman Correlation Coefficient
g
h
BP-R2
Relationship Strength
Bin Spearman Correlation
Negative Correlation
Positive Correlation

Figure 3


BP-R²
Relationship Strength
Bin Spearman Correlation
Negative Correlation
Positive Correlation

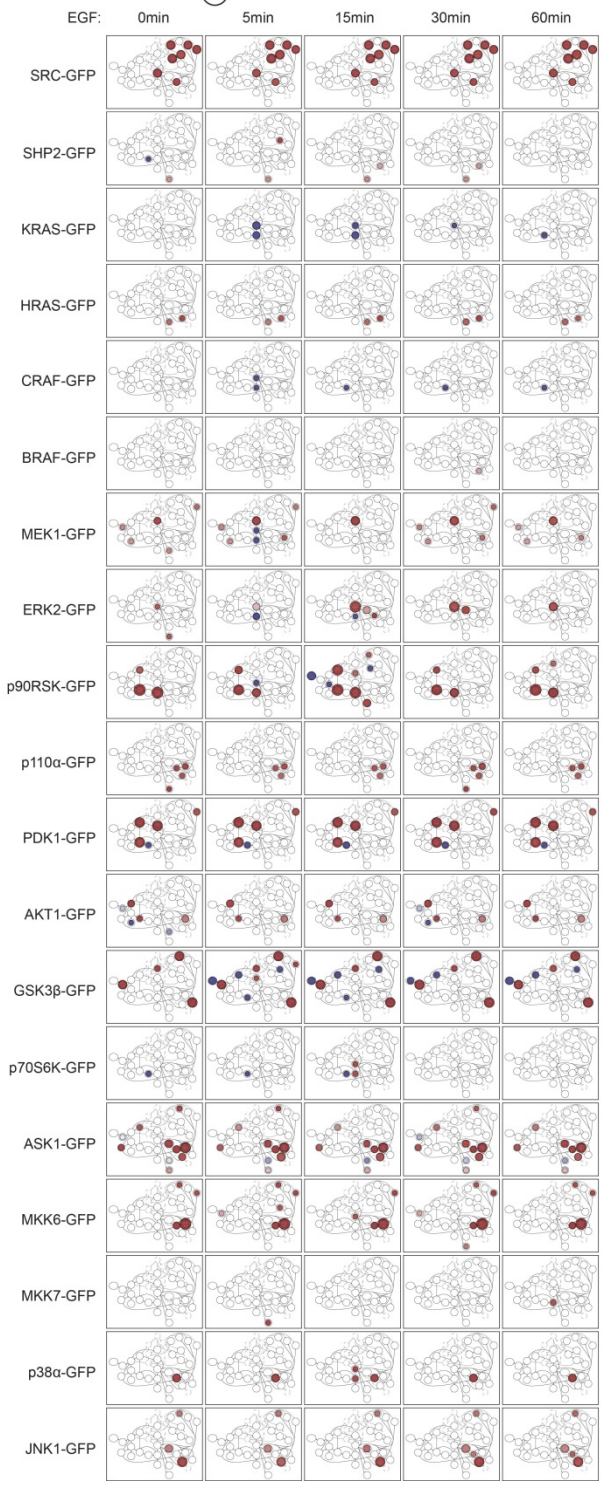
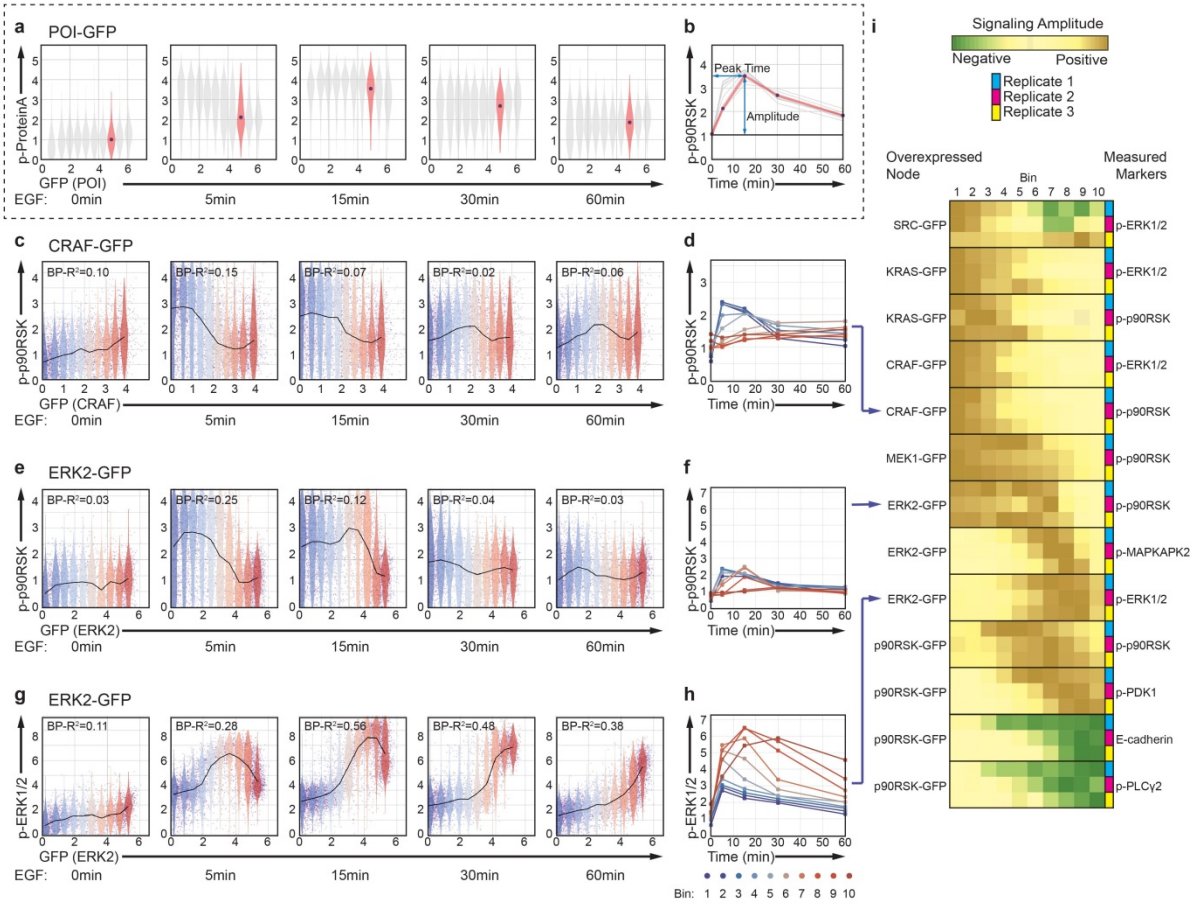
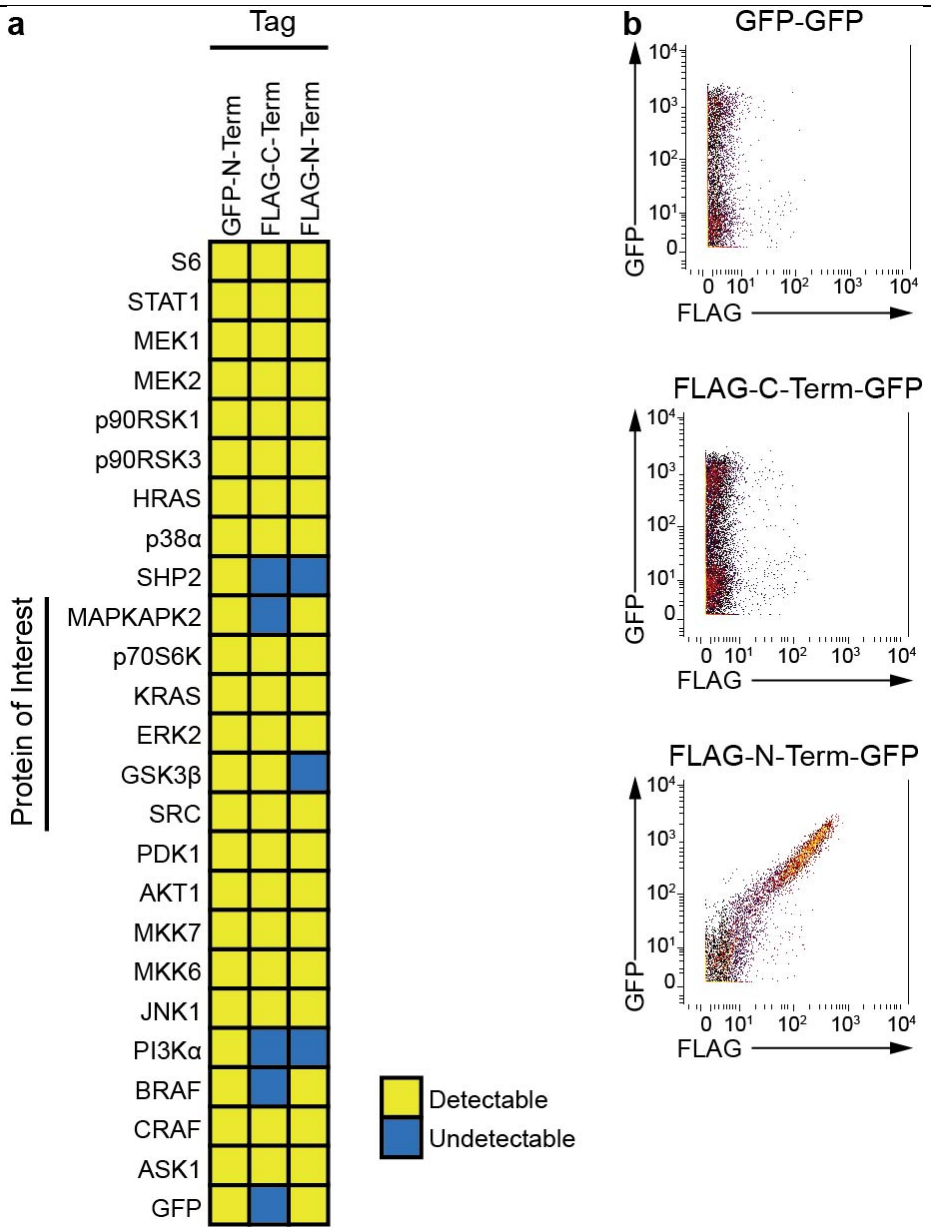

EGF:
0min
5min
15min
30min
60min
SRC-GFP
SHP2-GFP
KRAS-GFP
HRAS-GFP
CRAF-GFP
BRAF-GFP
MEK1-GFP
ERK2-GFP
p90RSK-GFP
p110α-GFP
PDK1-GFP
AKT1-GFP
GSK3β-GFP
p70S6K-GFP
ASK1-GFP
MKK6-GFP
MKK7-GFP
p38α-GFP
JNK1-GFP

Figure 4

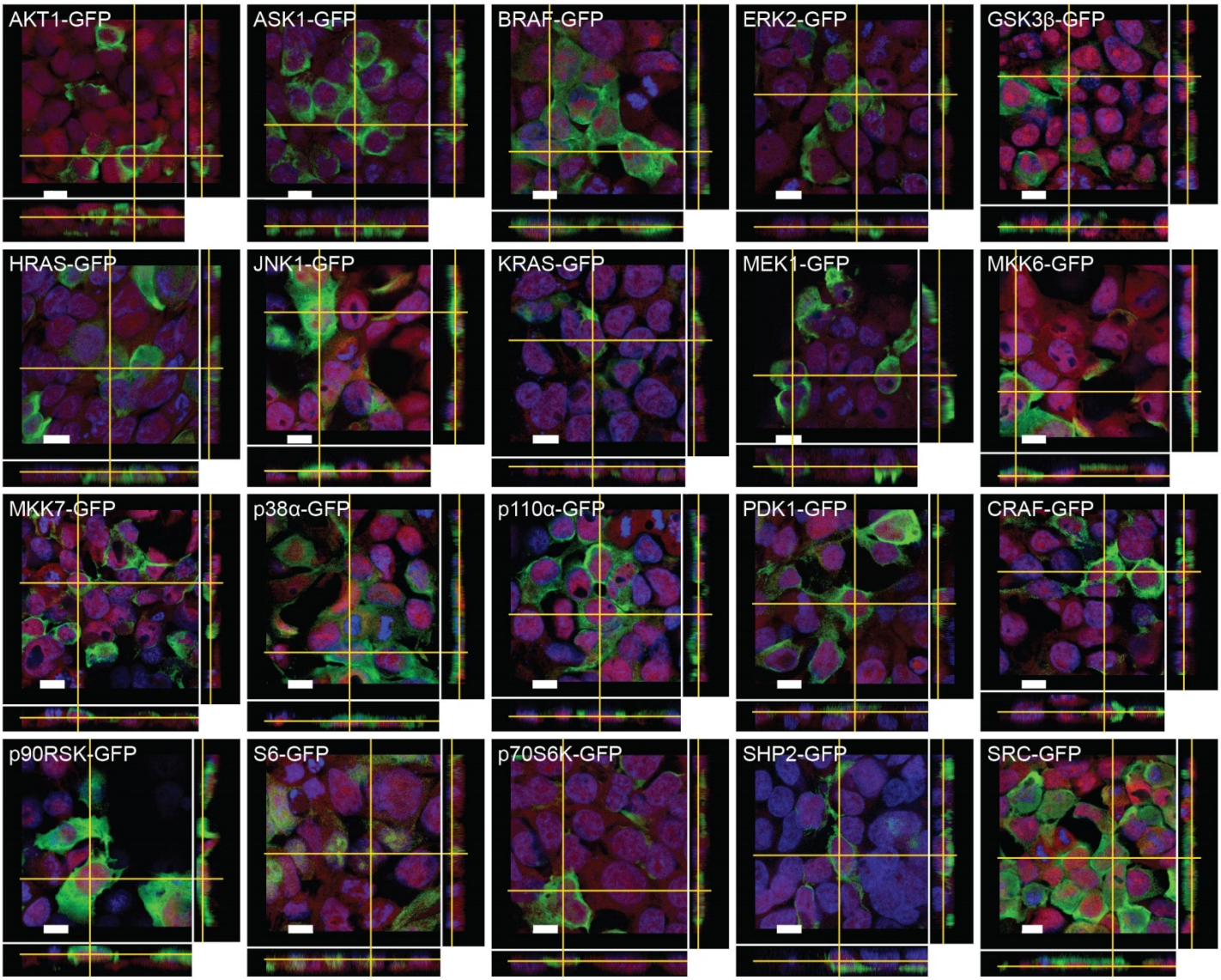

a
POI-GFP
p-ProteinA
GFP (POI)
EGF: 0min 5min 15min 30min 60min
b
Peak Time
Amplitude
p-p90RSK
Time (min)
c
CRAF-GFP
BP-R²=0.10
BP-R²=0.15
BP-R²=0.07
BP-R²=0.02
BP-R²=0.06
p-p90RSK
GFP (CRAF)
EGF: 0min 5min 15min 30min 60min
d
p-p90RSK
Time (min)
e
ERK2-GFP
BP-R²=0.03
BP-R²=0.25
BP-R²=0.12
BP-R²=0.04
BP-R²=0.03
p-p90RSK
GFP (ERK2)
EGF: 0min 5min 15min 30min 60min
f
p-p90RSK
Time (min)
g
ERK2-GFP
BP-R²=0.11
BP-R²=0.28
BP-R²=0.56
BP-R²=0.48
BP-R²=0.38
p-ERK1/2
GFP (ERK2)
EGF: 0min 5min 15min 30min 60min
h
p-ERK1/2
Time (min)
Bin: 1 2 3 4 5 6 7 8 9 10
i
Signaling Amplitude
Negative
Positive
Replicate 1
Replicate 2
Replicate 3
Overexpressed Node
Measured Markers
Bin
SRC-GFP
KRAS-GFP
KRAS-GFP
CRAF-GFP
CRAF-GFP
MEK1-GFP
ERK2-GFP
ERK2-GFP
ERK2-GFP
p90RSK-GFP
p90RSK-GFP
p90RSK-GFP
p90RSK-GFP
p-ERK1/2
p-ERK1/2
p-p90RSK
p-ERK1/2
p-p90RSK
p-p90RSK
p-p90RSK
p-MAPKAPK2
p-ERK1/2
p-p90RSK
p-PDK1
E-cadherin
p-PLCγ2




**Supplementary Figure 1**

Technique validation.

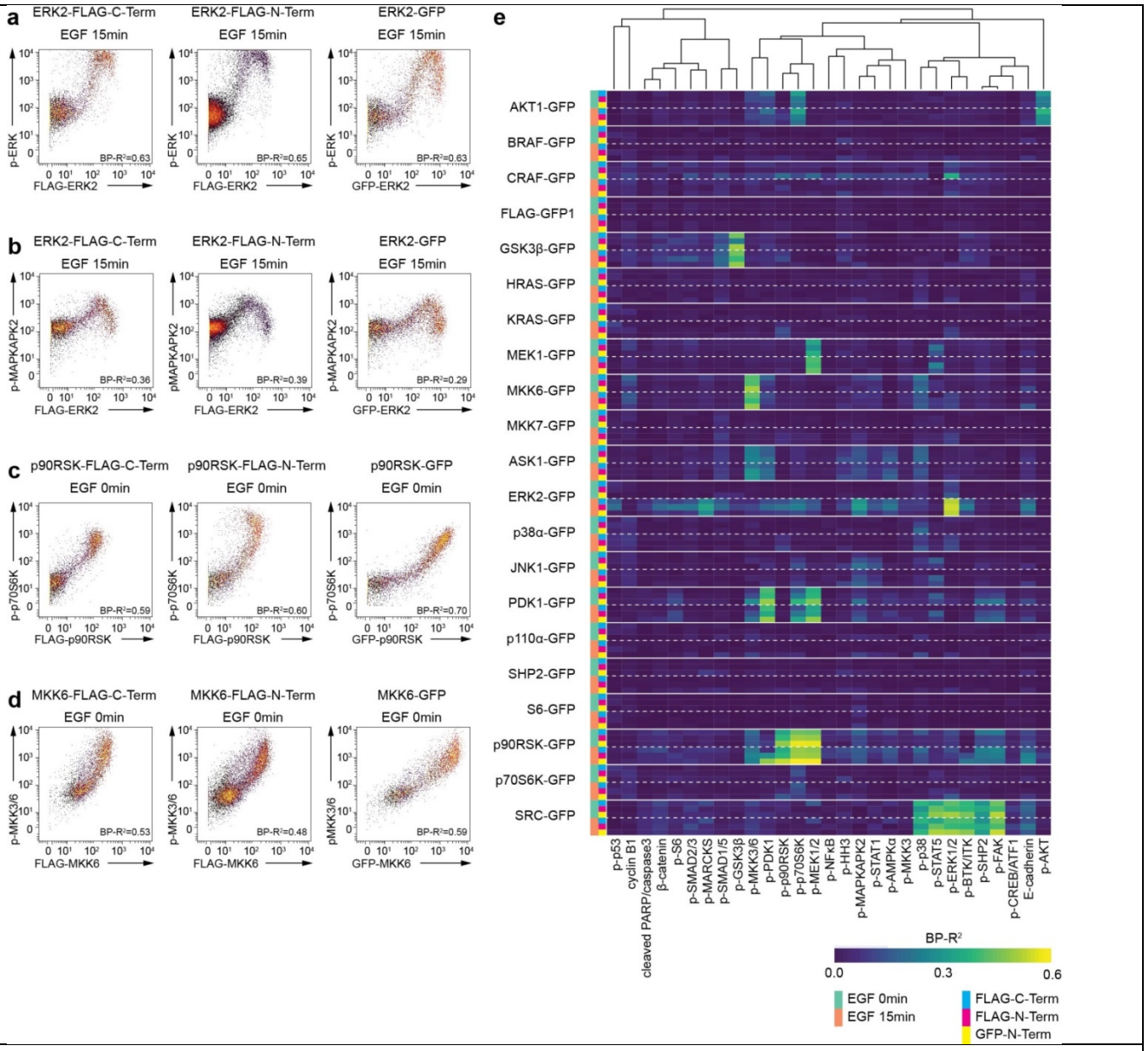
**(a)** Detection of GFP-N-terminal, FLAG-C-terminal, and FLAG-N-terminal tagged proteins. All GFP-tagged fusion proteins, but only 20 of the 25 FLAG-C-terminal tagged and only 22 of the 25 FLAG-N-terminal tagged proteins, were detected using mass cytometry. **(b)** HEK293T cells overexpressing GFP-GFP, FLAG-C-terminal-GFP, and FLAG-N-terminal-GFP fusion proteins were co-stained with anti-GFP and anti-FLAG antibodies. The fusion protein FLAG-C-terminal-GFP was detected by the anti-GFP antibody but not with the anti-FLAG antibody. This indicates that in certain contexts the FLAG tag is not accessible to the anti-FLAG antibody. The FLAG epitope may be masked due to protein folding or by the denaturation process that is part of our experimental protocol.




**Supplementary Figure 2**

GFP-tagged POIs have normal localization.

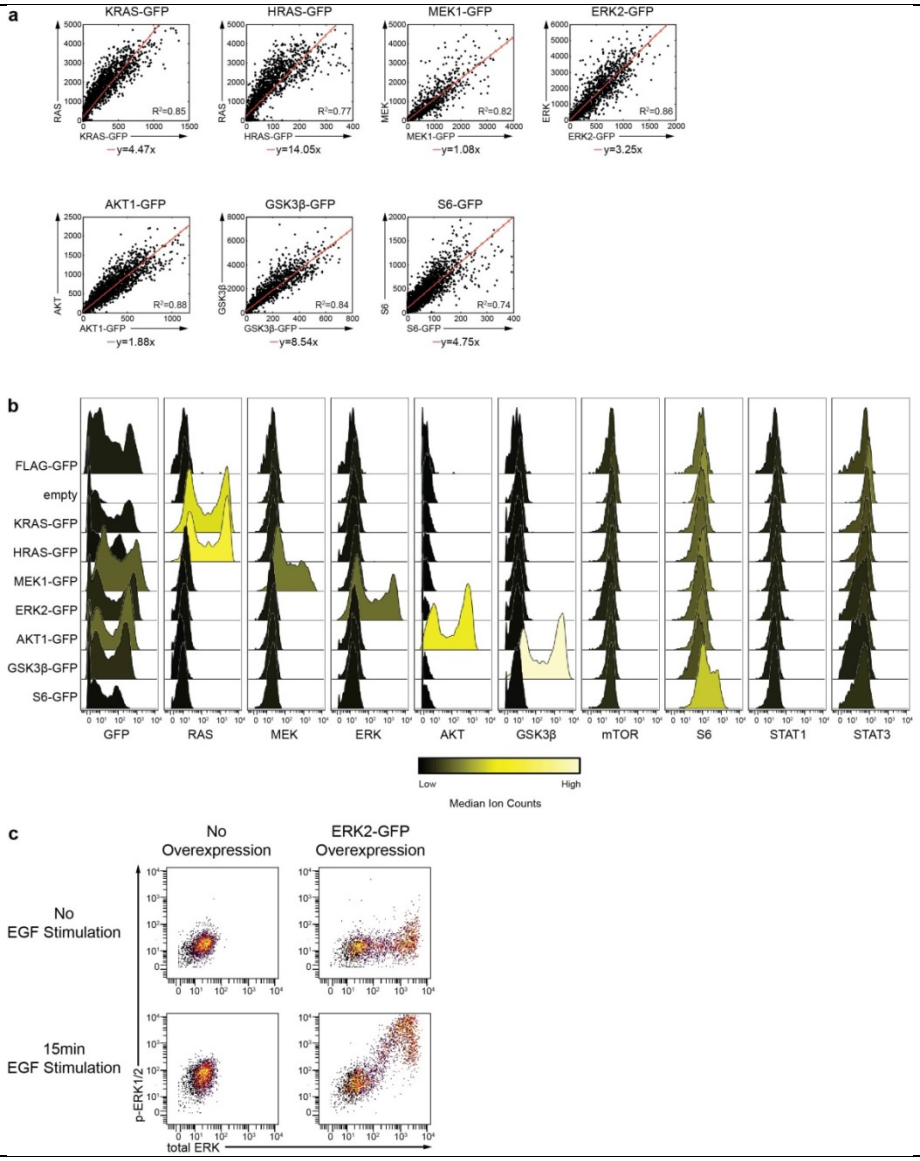
HEK293T cells that overexpressed the GFP-tagged POIs used in this study were imaged with confocal microscopy. For each POI, the main panel shows the image in a given z-depth; the bottom panel and the side panel shows x-z and y-z cross-sectional images, respectively. POI-GFP subcellular localization was determined by overlapping with two control stains: Hoechst 33343 for the nucleus and Alexa Fluor 647 carboxylic acid succinimidyl ester indicating the cell outline. The POI-GFP localization was verified by comparison with information of the UniProt subcellular localization database (Supplementary Table 4).




**Supplementary Figure 3**

GFP tag does not disrupt catalytic activities of POIs.

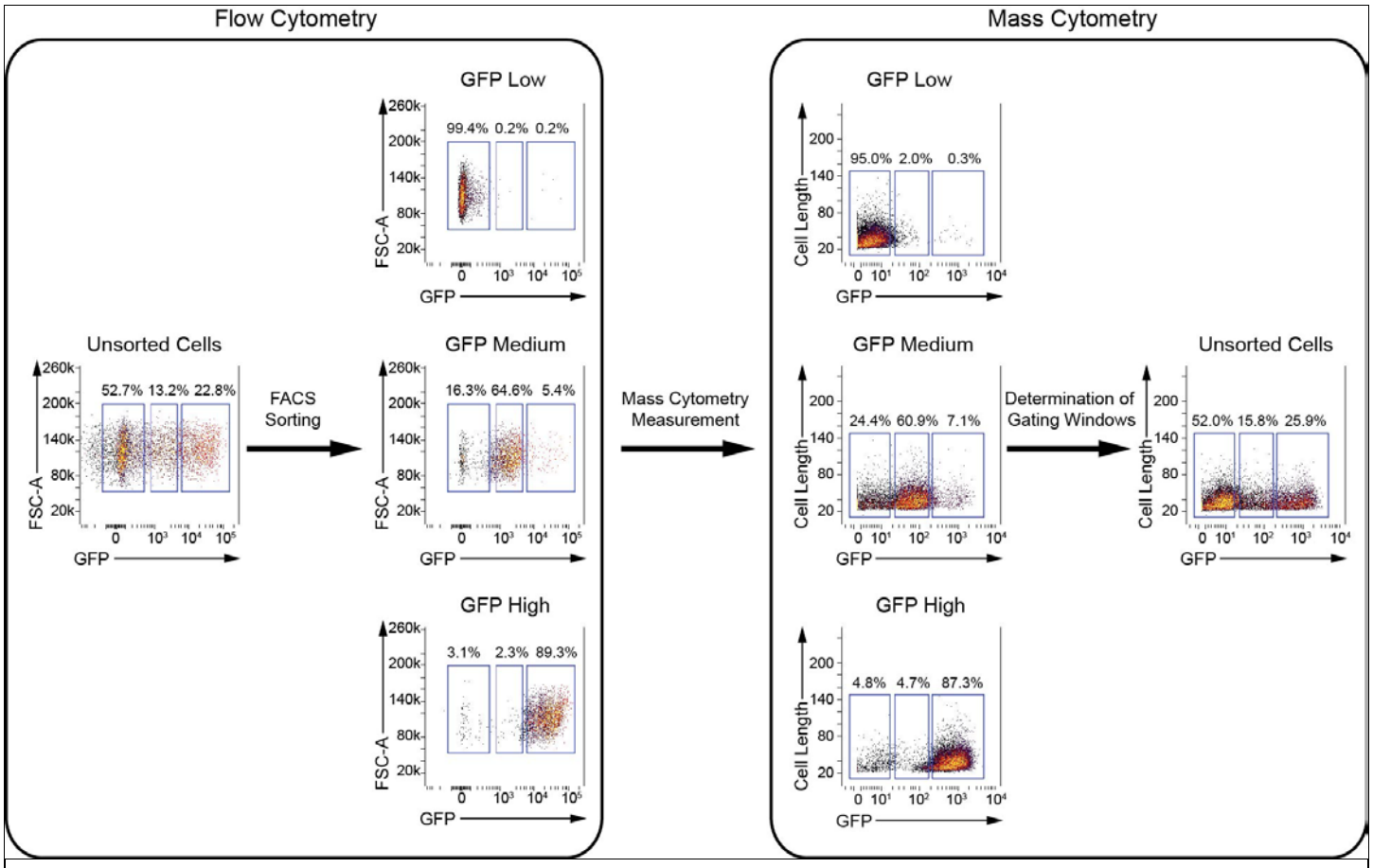
**(a-d)** Catalytic activities of GFP-tagged POIs were compared with FLAG-C-terminal and FLAG-N-terminal tagged POIs. The examples shown here indicate that the GFP tag did not alter signaling relationships or signaling dynamics after EGF stimulation (the complete dataset with comparison of all constructs used in this study is shown in Supplementary File 1). **(e)** Heat map showing abundance-dependent signaling relationship strengths from overexpressed POIs with three different tags as determined by BP-$R^2$ analysis (Supplementary Figure 11 and Methods). Measured markers showing at least one strong relationship in any of the conditions were included in the heat map. Strong relationships were detected independently of tag. BP-$R^2$ values slightly vary for the 3 tags, due to the antibody accessibility and differences in transfection efficiencies (Supplementary Figure 1).




**Supplementary Figure 4**

Total protein antibody staining of HEK293T cells overexpressing a GFP-tagged POI.

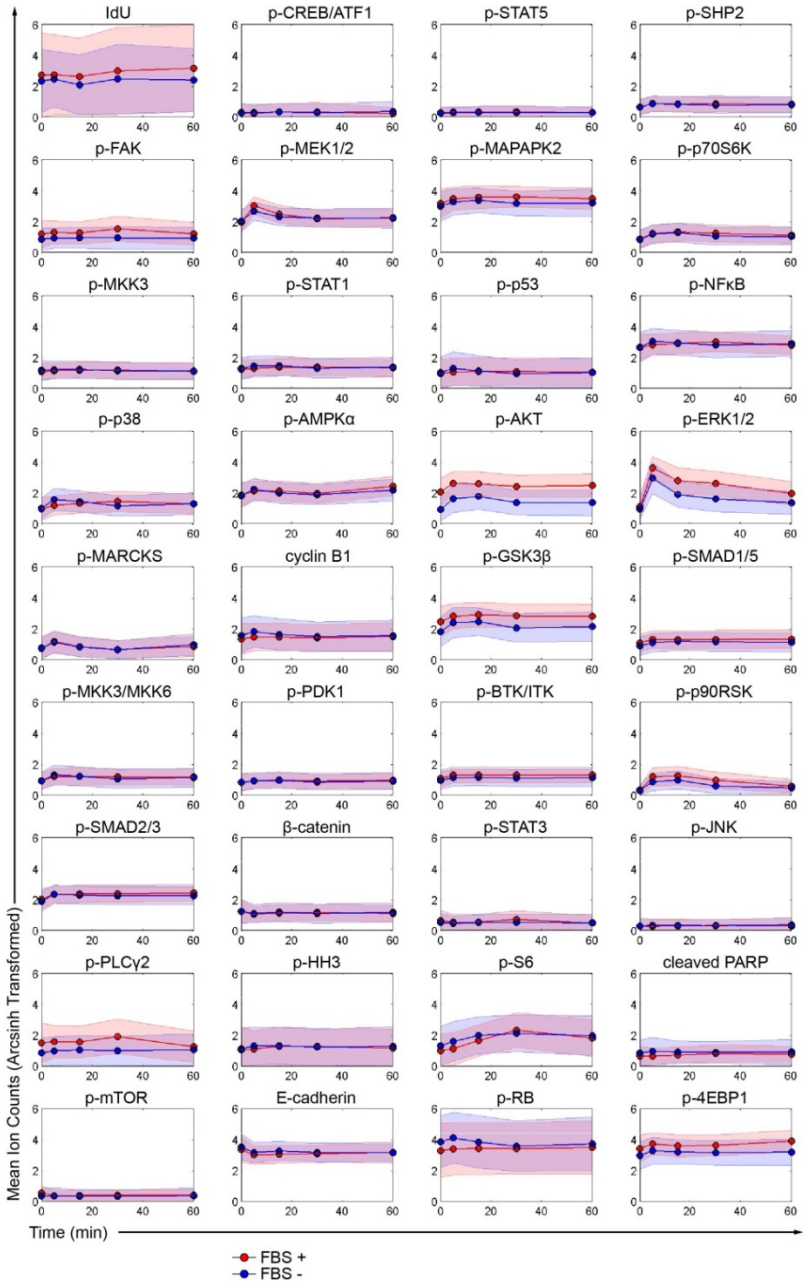
**(a)** HEK293T cells transfected with KRAS-GFP, HRAS-GFP, MEK1-GFP, ERK2-GFP, AKT1-GFP, GSK3 β-GFP, or S6-GFP for 18 h were stained with anti-total POI and anti-GFP antibodies. A linear regression analysis for each pair was performed in the original scale. R2 ranges from 0.74 to 0.88, indicating the total POI is linearly correlated with GFP and that the POI overexpression does not alter the expression of the endogenous POI. **(b)** The same cells were stained with nine antibodies to quantify total protein as well as with a GFP antibody. Median ion counts for all measured markers are shown. Overexpression of a POI-GFP for 18 h does not cause notable changes in the measured network nodes. **(c)** ERK2-GFP transfected HEK293T cells and the untransfected control with or without EGF stimulation were stained for total-ERK and phospho-ERK (Thr202/Tyr204). The dynamic range in the overexpression condition allows observation of abundance-dependent signaling relationships. With total ERK staining, the same signaling relationships as shown in the Supplementary Figure 2b is recapitulated, verifying GFP as an indicator of POI expression level.




**Supplementary Figure 5**

Comparison of mass cytometry and flow cytometry (FACS).

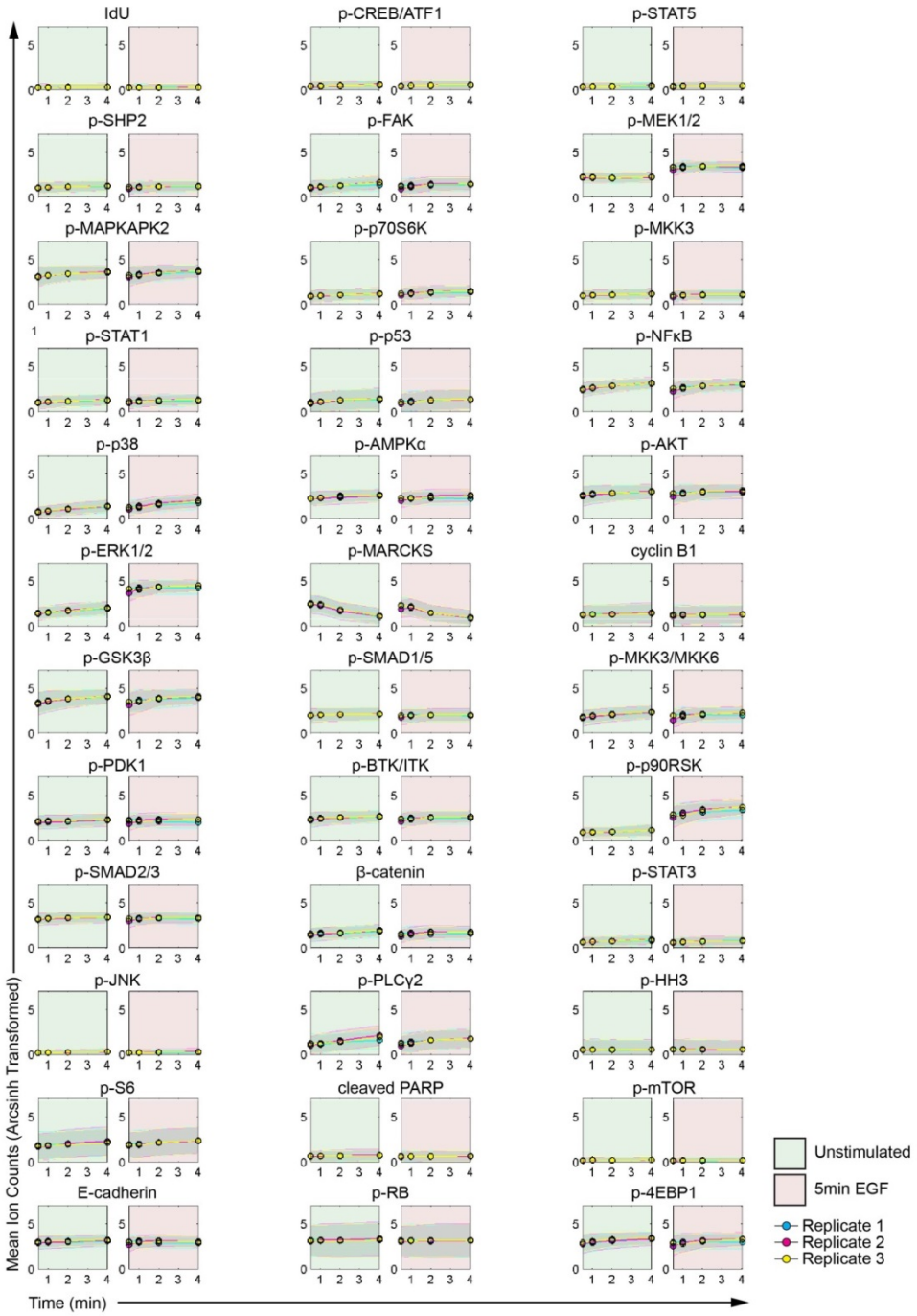
HEK293T cells were transfected with the FLAG-GFP overexpression vector. With flow cytometry, cells were gated into GFP low, medium, and high populations with the gating strategy shown in the left panel. With mass cytometry, each of the three sorted populations was measured independently to determine the gating windows. Unsorted cells were then assessed by the mass cytometry. The maximum difference in population percentage between mass cytometry and flow cytometry was less than 3%.




**Supplementary Figure 6**

Comparison of EGF stimulations in starved (FBS is absent) and non-starved (FBS is present) cell culture conditions.

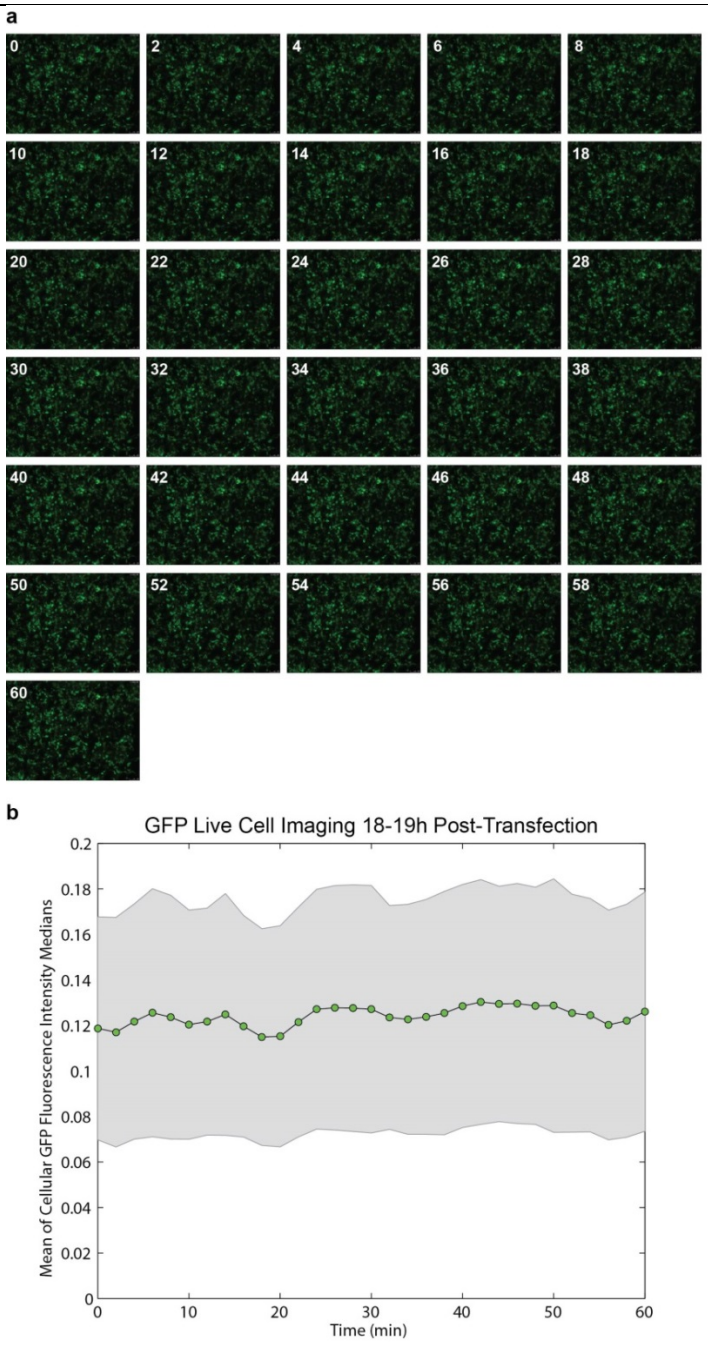
HEK293T cells were stimulated with EGF with or without FBS over a 1-h time course. In the non-starved condition basal signaling states of the major MAPK/ERK or AKT pathway components were higher than in starved conditions, but these elevated levels did not affect the signaling responses to the EGF stimulation. Mean value of each sample is shown with circle. Standard deviation is indicated by shaded area.




**Supplementary Figure 7**

TrypLE treatment time course.

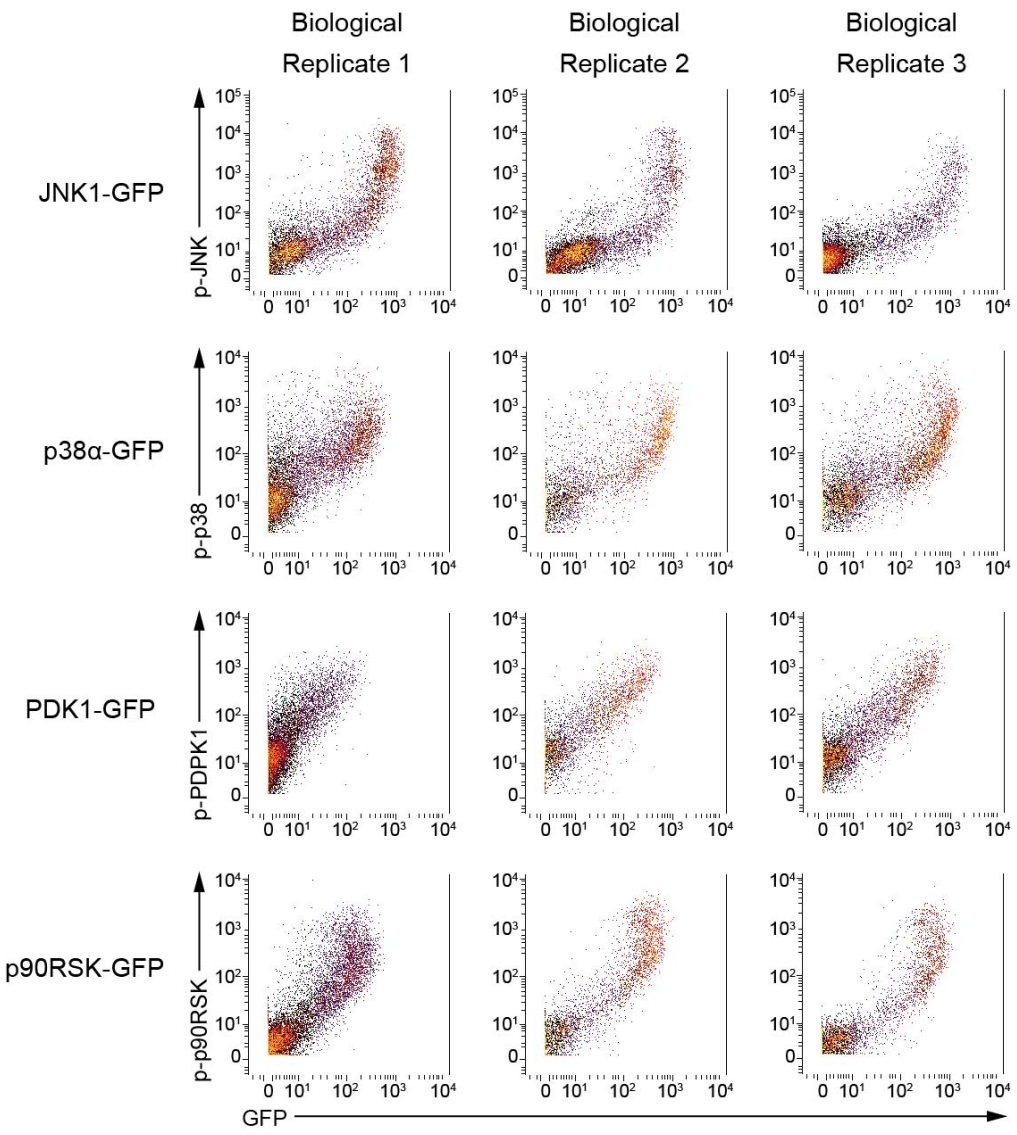
HEK293T cells were treated with TrypLE for 30 s, 1 min, 2 min, or 4 min with or without EGF stimulation for 5 min (time from EGF addition to PFA crosslinking). Within the first 2-min TrypLE treatment (i.e., the time after which we quenched cells in all experiments), only phosphorylation of Ser167/170 on MARCKS varied relatively. Mean value of each sample is shown with circle. Standard deviation is indicated by shaded area.




**Supplementary Figure 8**

Live imaging of GFP fluorescence at 18 to 19 h after HEK293T cells were transfected with a FLAG-GFP construct.

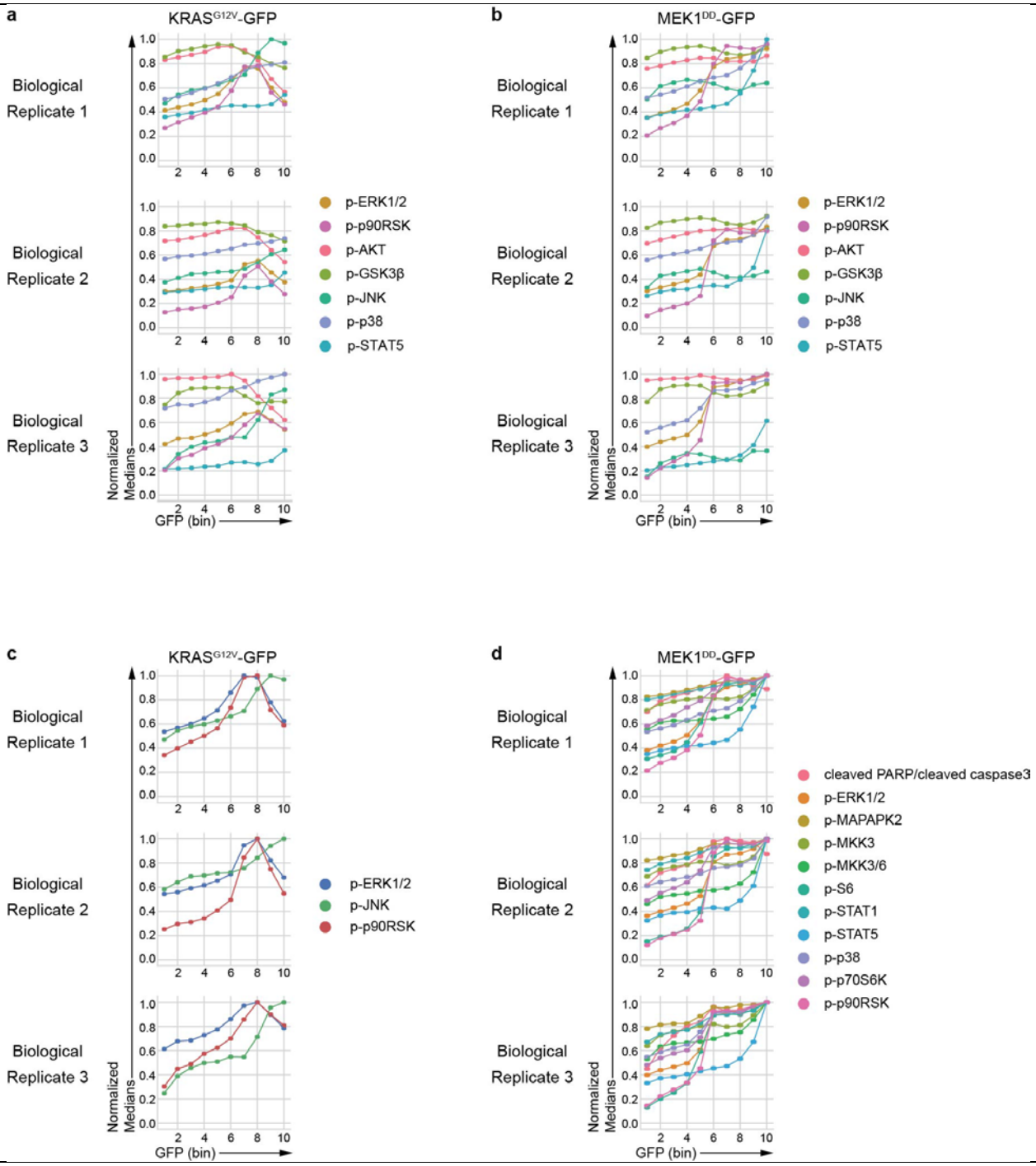
Quantification of the GFP intensity showed a slight increase of 5.4% over the 1-hr time course. There was a fluctuation in total GFP signal, indicating that the 5.4% increase is most likely attributable to technical variability of the measurement. The analysis of signaling relationships in our study was performed based on a binning strategy on arcsinh transformed GFP ion counts (mass cytometry). Thus, the measured change will not significantly affect the binning over the time course. Standard deviation is indicated by shaded area.




**Supplementary Figure 9**

Reproducibility test.

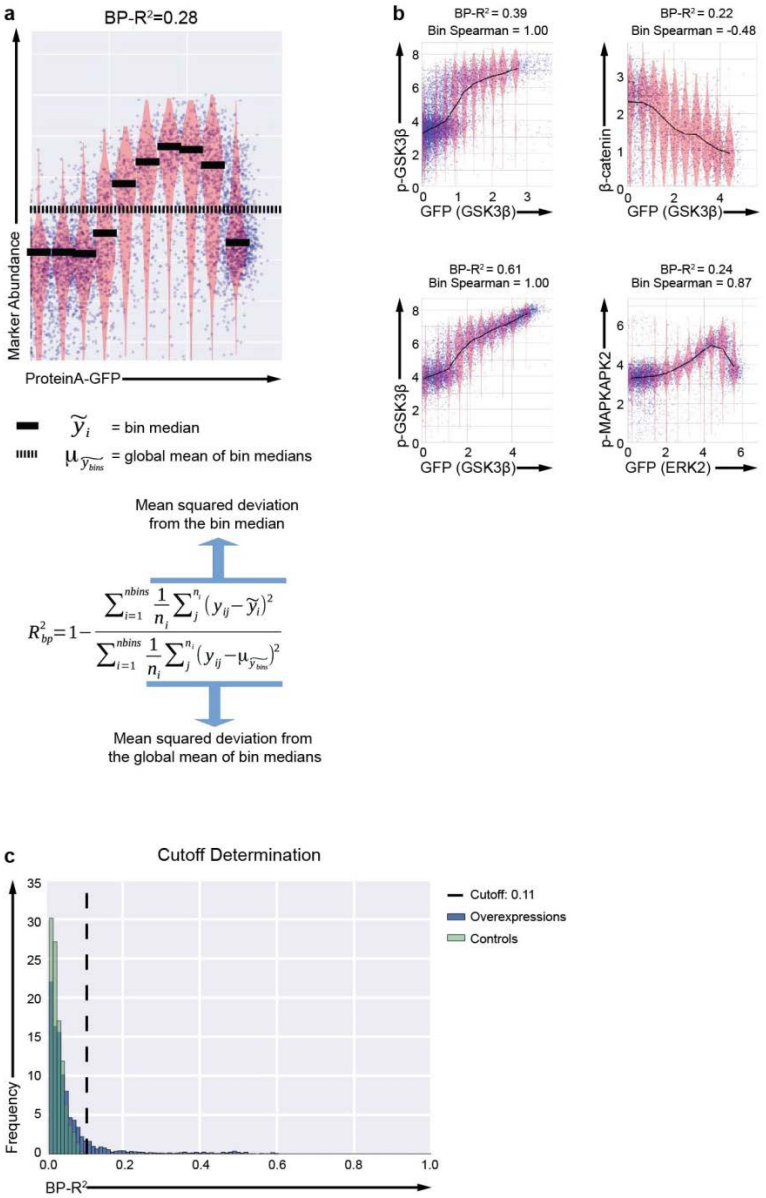
Different batches of HEK293T cells were transfected with JNK1-GFP, P38α-GFP, PDK1-GFP, or p90RSK-GFP constructs, stained, and analyzed by mass cytometry on three different days. Highly consistent signaling responses were observed among the three biological replicates.




**Supplementary Figure 10**

Abundance-dependent signaling analysis performed in replicates for the shown mutants.

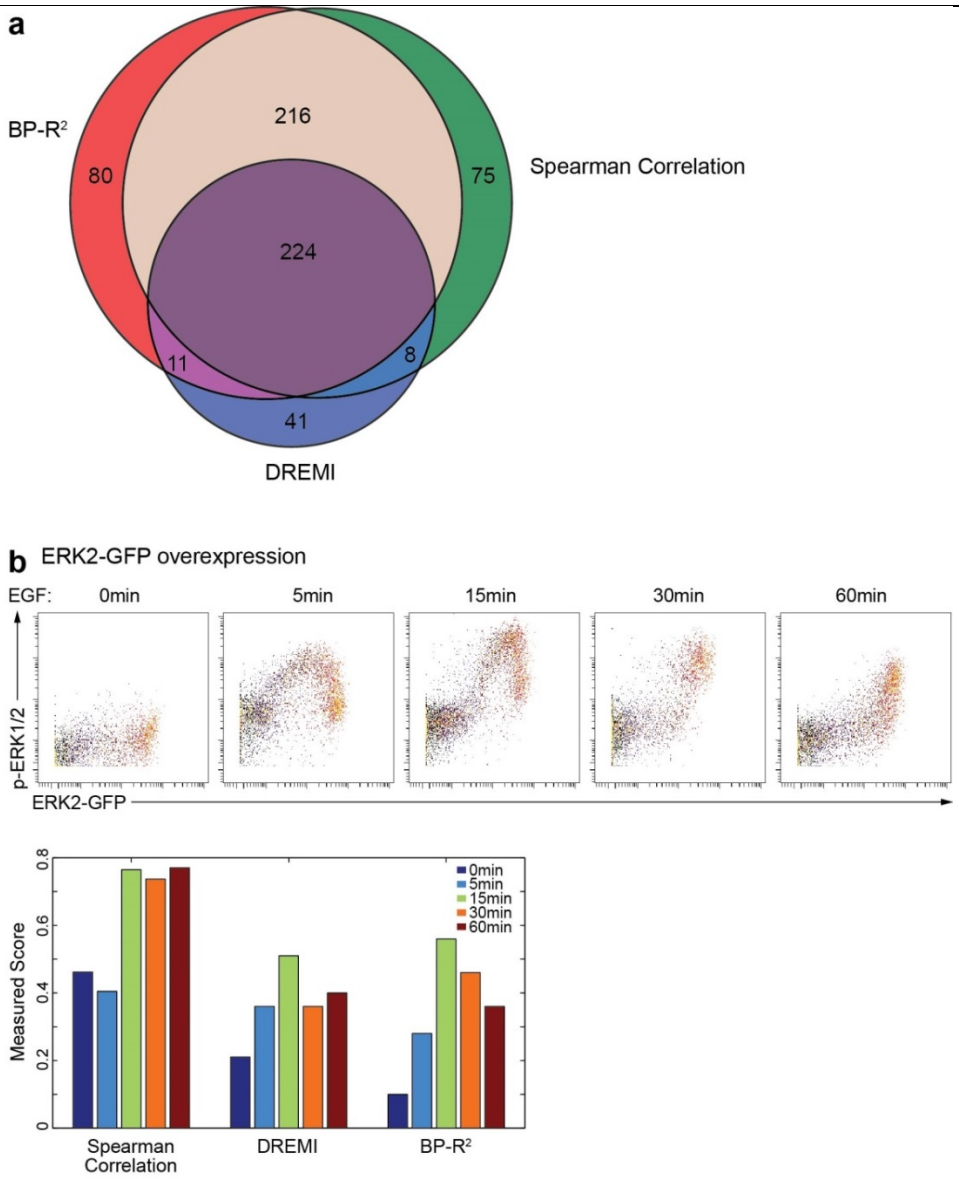
Panels **(a)** and **(b)** show analyses of representative phosphorylation sites in the MAPK/ERK, AKT, stress pathways, and the STAT5 protein in cells in which **(a)** KRAS$^{G12V}$-GFP and **(b)** MEK1$^{DD}$-GFP was overexpressed. Panels **(c)** and **(d)** show all relationships that passed the BP-$R^2$ threshold (see Methods for details) for the **(c)** KRAS$^{G12V}$-GFP and **(d)** MEK1$^{DD}$-GFP overexpression experiments.




**Supplementary Figure 11**

Binned pseudo $R^2$ (BP-$R^2$) analysis.

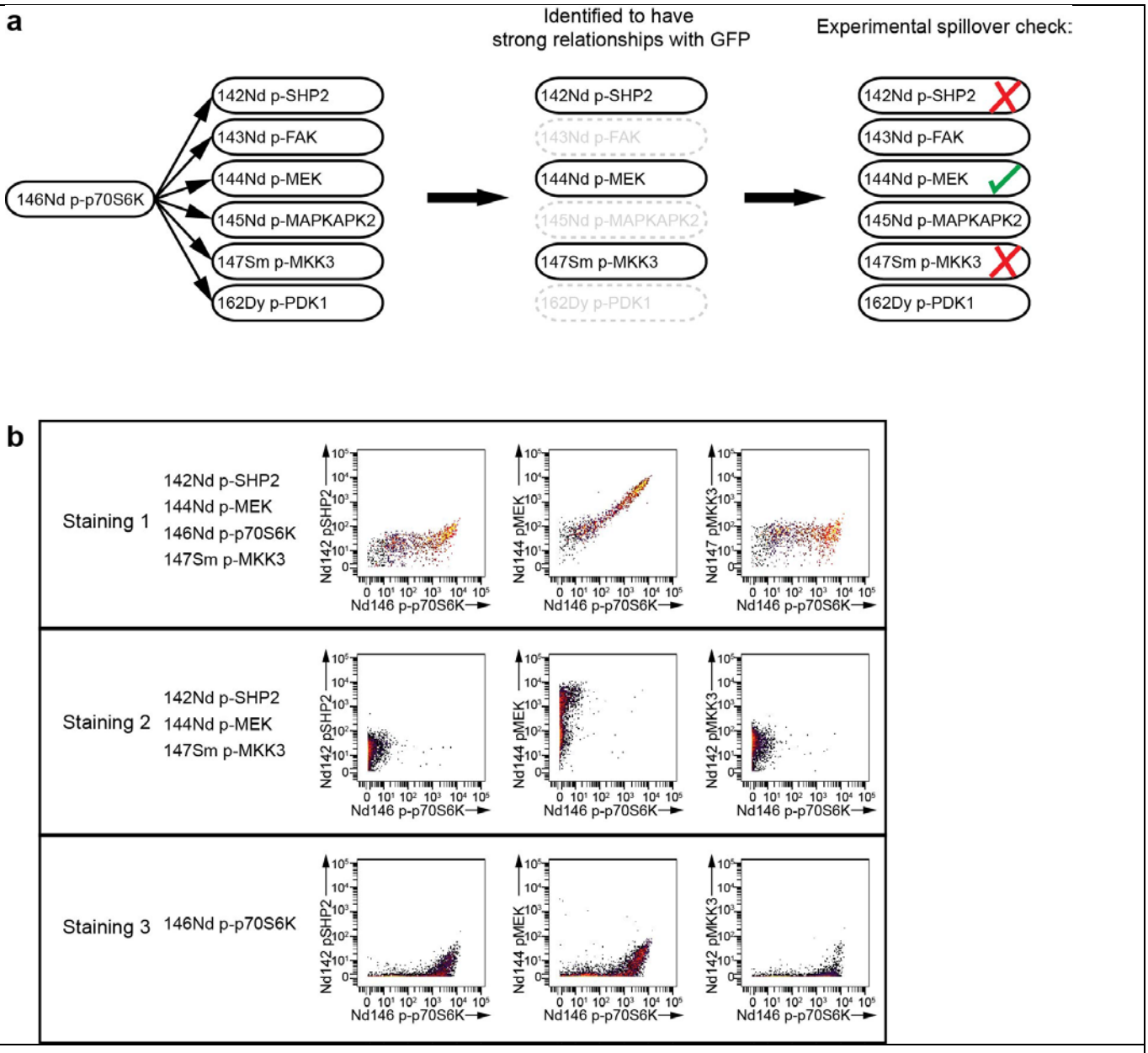
**(a)** BP-$R^2$ analysis considers deviation from bin median versus the global mean of bin medians. **(b)** Examples of BP-$R^2$ and Spearman correlation of bin medians values. The top left and top right plots show examples of positive and negative Spearman correlations of bin medians. The top left and bottom left plots show replicates of the same overexpression condition and how a (supposedly) increased noisiness affects the BP-$R^2$ values. The bottom right plot shows a complex signaling relationship with the corresponding BP-$R^2$ value. The BP-$R^2$ metric detects complex arbitrary relationship (bottom right). **(c)** Density distribution of the median BP-$R^2$ values for the 700 POI-GFP-marker relationships from the negative controls (FLAG-GFP, untransfected) and the 3500 POI-GFP-marker relationships of the signaling node overexpression conditions. Cutoff for strong signaling relationships were determined at a median BP-$R^2$ value of 0.11, the highest median BP-$R^2$ of the negative controls.




**Supplementary Figure 12**

Benchmark of BP-$R^2$ against other methods used to identify relationships in mass cytometry data.

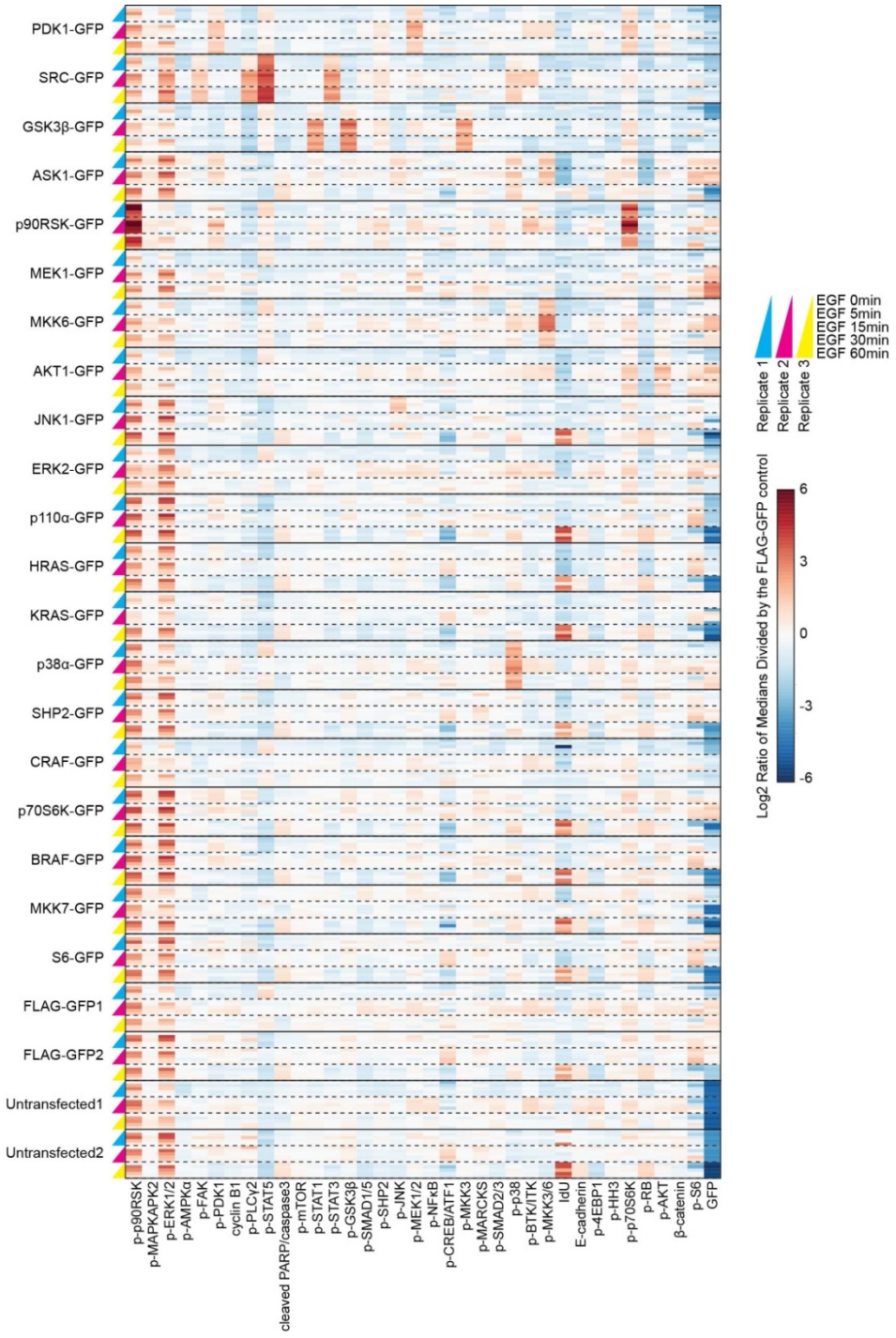
**(a)** Venn diagram of strong relationships detected by BP-$R^2$, Spearman correlation, and DREMI in our dataset using the same cutoff - the 99 percentile of the BP-$R^2$ / Spearman correlation / DREMI score in the control groups (FLAG-GFP overexpression and the untransfected cells). BP-$R^2$ outperforms the other two measures. **(b)** BP-$R^2$, Spearman correlation, and DREMI measurements of signaling relationship strength between p-ERK1/2 and overexpressed ERK2-GFP. BP-$R^2$ is suitable for analyzing non-monotonic signaling relationships and outperforms the other two measures in representing actual signaling activation status.




**Supplementary Figure 13**

Analysis of signal spill over among mass channels.

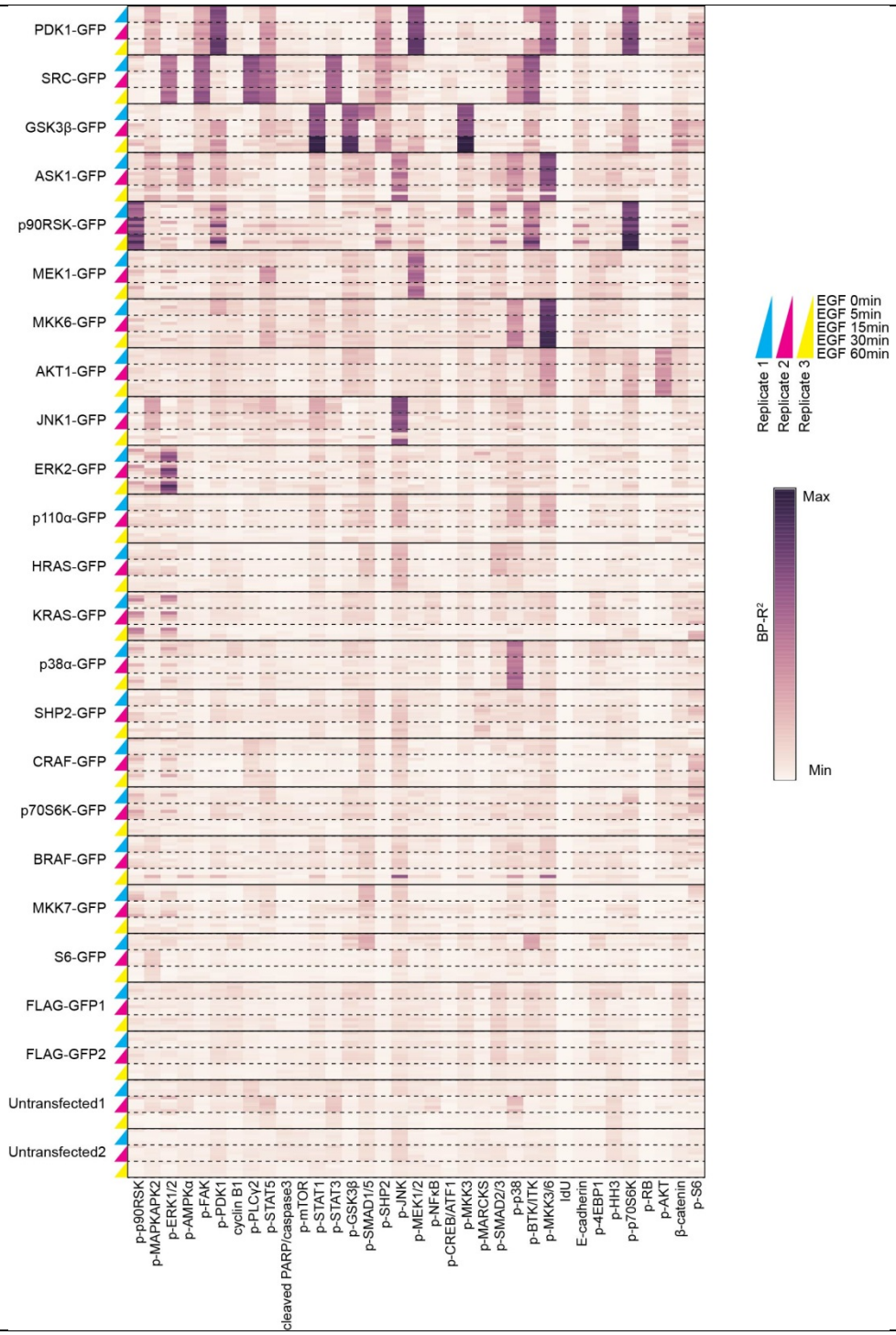
**(a)** Strategy to exclude spill over among mass channels. When strong signaling relationships as determined by BP-$R^2$ were identified (measured phosphorylation of p70S6K in the p90RSK-GFP overexpression is shown here as a selected example), all other potentially affected channels (details in Methods) were evaluated for spillover that might have led to a high BP-$R^2$ value. Using an experimental spillover filter **(b)**, spillover-affected relationships were discarded. Here three groups of antibody stains were performed simultaneously: First, all antibodies; second, all antibodies except for the one that potentially causes spillover; third, only the antibody that potentially causes spillover. If spillover induced background was over 10% of the actual ion counts, the channel was discarded from the analysis.




**Supplementary Figure 14**

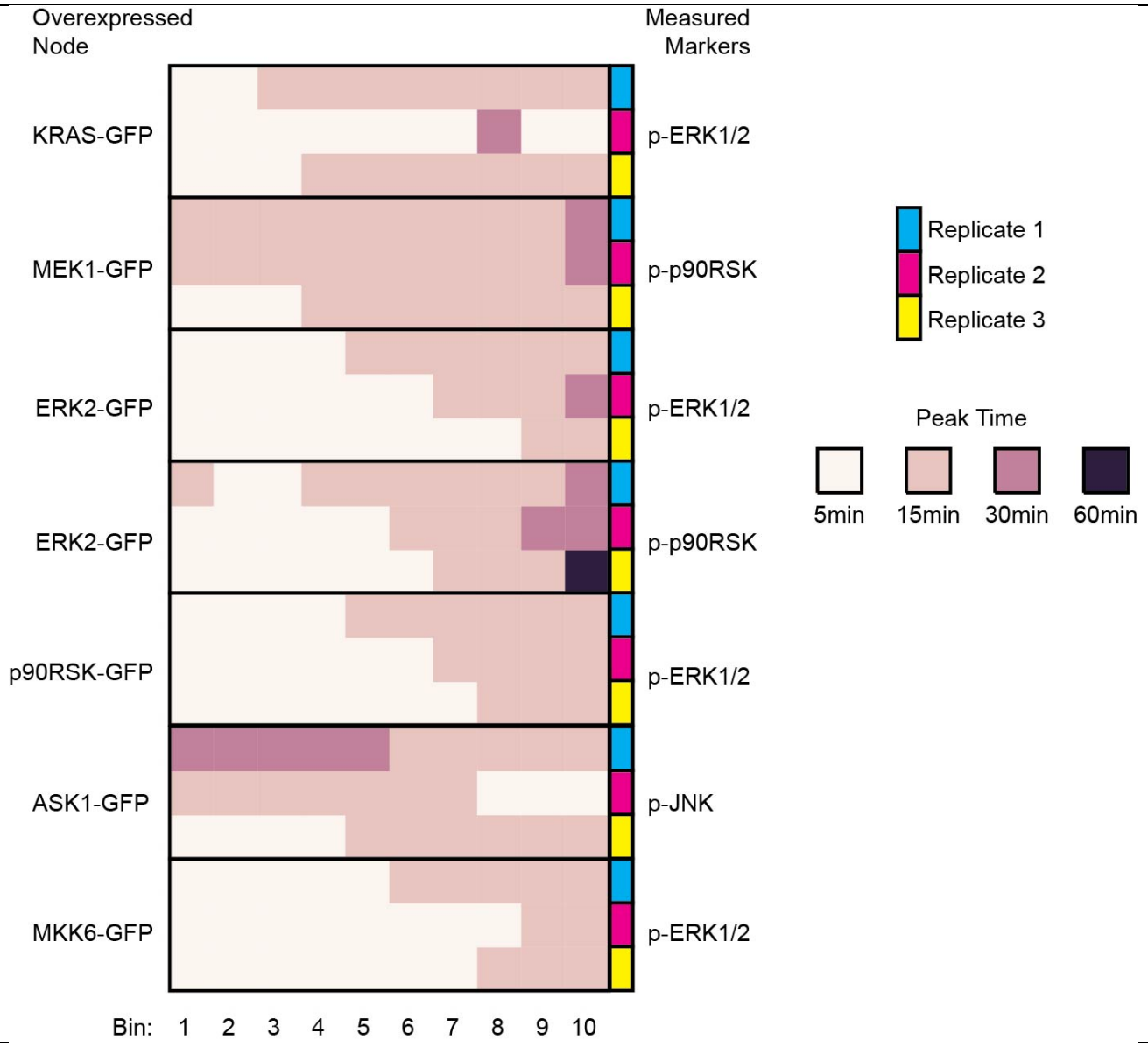
Heat map of median intensities of all measured markers at 0, 5, 15, 30, and 60 min post-EGF stimulation in all overexpression conditions (Table 1).

Data visualized as log2 of the ratio of the median signals divided by the mean of median signals of the FLAG-GFP controls at time point 0.



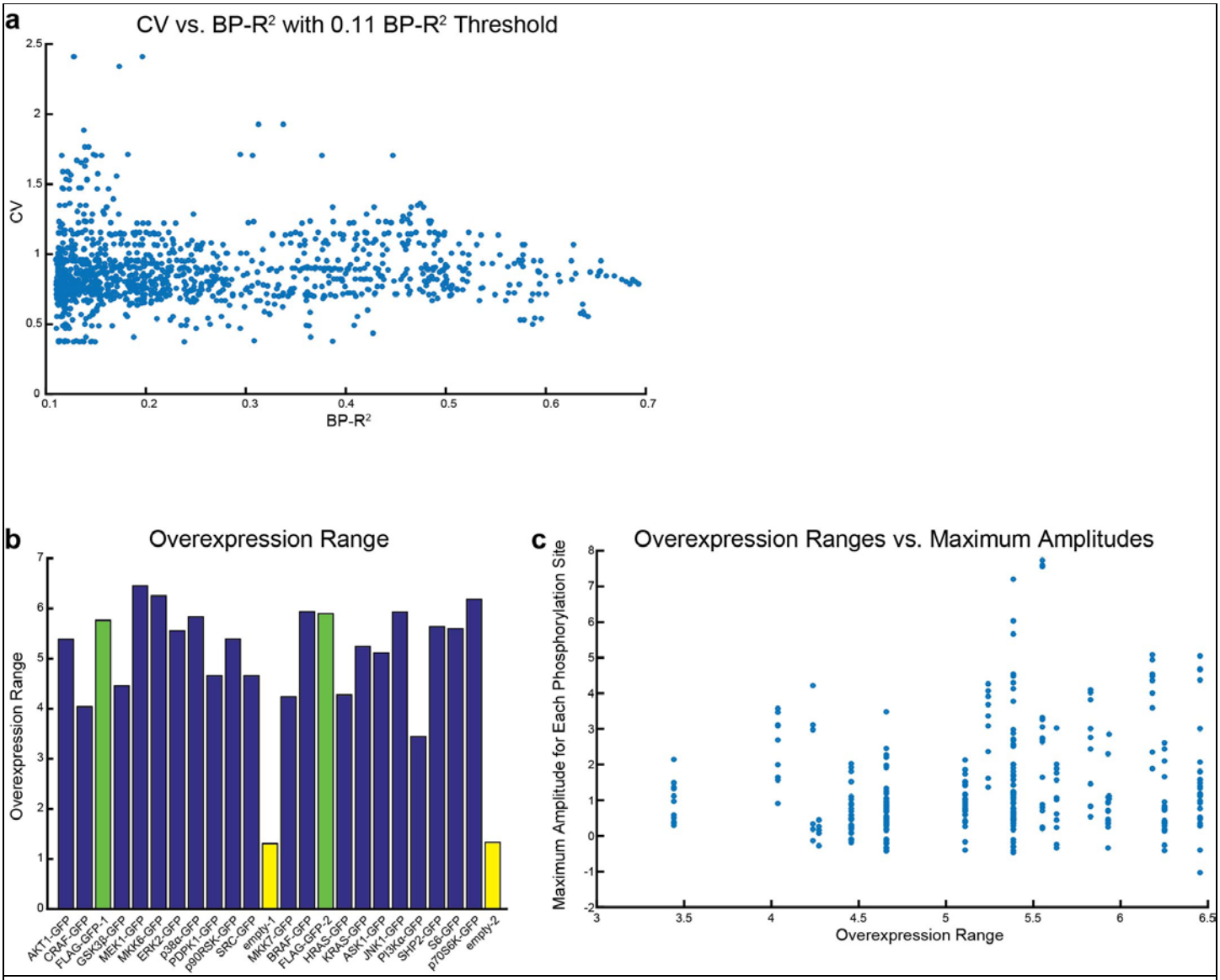

**Supplementary Figure 15**

Heat map of BP-$R^2$ values of all measured markers versus GFP signals at 0, 5, 15, 30, and 60 min post-EGF stimulation in all overexpression conditions.




**Supplementary Figure 16**

Heat map of consistent and robust examples for overexpression-induced phosphorylation site abundance peak time changes after EGF stimulation for each of the three replicates.




**Supplementary Figure 17**

Post-transcriptional constraint analysis of overexpressed POIs.

**(a)** Coefficient of variation (CV) was computed for each strong signaling relationship that had a BP-$R^2$ value above 0.11 (i.e., a strong signaling relationship), and CVs were plotted against BP-$R^2$. No correlation was observed. **(b)** Overexpression ranges (median value of GFP in Bin10 minus the median value of GFP in Bin1) calculated for all POIs. **(c)** Maximum amplitudes of phosphorylation sites were independent of the level of overexpression of the POIs.

## Result 2: Analysis of the human kinome and phosphatome reveals diseased signaling networks induced by overexpression

Xiao-Kang Lun[1,2], Damian Szklarczyk[1]*, Attila Gábor[3]*, Nadine Dobberstein[1], Vito RT Zanotelli[1,4], Julio Saez-Rodriguez[3, 5#], Christian von Mering[1#], and Bernd Bodenmiller[1]

[1] Institute of Molecular Life Sciences, University of Zürich, Zürich, Switzerland

[2] Molecular Life Science Ph.D. Program, Life Science Zürich Graduate School, ETH Zürich and University of Zürich, Zürich, Switzerland

[3] Joint Research Centre for Computational Biomedicine, Faculty of Medicine, RWTH Aachen University, Aachen, Germany

[4] Systems Biology Ph.D. Program, Life Science Zürich Graduate School, ETH Zürich and University of Zürich, Zürich, Switzerland

[5]European Bioinformatics Institute, European Molecular Biology Laboratory (EMBL-EBI) Hinxton, Cambridge, UK

*These authors contributed equally to the manuscript.

[#] These authors contributed equally to the manuscript.

Correspondence should be addressed to

B.B. (bernd.bodenmiller@imls.uzh.ch)

## Abstract

Kinase and phosphatase overexpression drives tumorigenesis and drug resistance in many cancer types. Signaling networks reprogrammed by protein overexpression remain largely uncharacterized, hindering discovery of paths to therapeutic intervention. We previously developed a single cell proteomics approach based on mass cytometry that enables quantitative assessment of overexpression effects on the signaling network. Here we applied this approach in a human kinome- and phosphatome-wide study to assess how 649 individually overexpressed proteins modulate the cancer-related signaling network in HEK293T cells. Based on these data we expanded the functional classification of human kinases and phosphatases and detected 208 novel signaling relationships. In the signaling dynamics analysis, we showed that increased ERK-specific phosphatases sustained proliferative signaling, and using a novel combinatorial overexpression approach, we confirmed this phosphatase-driven mechanism of cancer progression. Finally, we identified 54 proteins that caused ligand-independent ERK activation with potential as biomarkers for drug resistance in cells carrying BRAF mutations.

## Introduction

Kinases and phosphatases control the reversible process of phosphorylation, which regulates protein structure, activity, and localization. Kinases and phosphatases are organized as signaling networks that compute extracellular signals into transcriptional, functional, and phenotypical responses. Deregulation of signaling networks can lead to the initiation and progression of many types of human disease including cancer (Fleuren et al., 2016; Julien et al., 2007), consequently they are a focal point of life science research. Kinases and phosphatases have been classified based on genomic and proteomic analyses (Chen et al., 2017; Manning et al., 2002; Sacco et al., 2012a). Signaling network structure has been studied by mapping physical interactions of kinases and phosphatases in steady and dynamic states using biochemical approaches and reporter assays in yeast and human cells (Barrios-Rodiles et al., 2005; Breitkreutz et al., 2010; Couzens et al., 2013; Horn et al., 2011). Using *in vitro* kinase assays and motif-based predictions, the specificity and targets of many kinases have been revealed (Linding et al., 2007; Mok et al., 2010; Yu et al., 2009). Kinase and phosphatase perturbations have also been applied to systematically determine phosphorylation abundance changes in yeast and human cells (Bodenmiller et al., 2010; Ochoa et al., 2016; Sacco et al., 2012b). A systematic analysis of how each human kinase and phosphatase modulates signaling network structure and dynamics, however, so far is absent.

Mutation-induced signaling network rewiring and modulation of signaling dynamics have been systematically characterized (Creixell et al., 2015; Pawson and Warner, 2007), providing a basis for identification of targeted therapies in cancer (Hennessy et al., 2005; Logue and Morrison, 2012). Independently of mutations, kinase overexpression drives tumorigenesis in multiple cancer types and is a critical factor in drug resistance (Eralp et al., 2008; Santarius et al., 2010; Shaffer et al., 2017). Overexpression of phosphatases has been shown to mediate cancer progression and has been linked to the poor prognosis of patients (Julien et al., 2011; Liu et al., 2016; De Vriendt et al., 2013). Overexpression-induced signaling modulation remains largely uncharacterized because factors such as genetic instability induce highly heterogeneous quantities of deregulated signaling proteins in cancer (Abbas et al., 2013), making conventional cell population-based analysis inapplicable. Only recently have technologies emerged that account for such heterogeneity, and can comprehensively quantify signaling network behavior with single-cell resolution. This resolution is required to characterize the abundance-related cellular signaling states and phenotypical alterations caused by a given kinase and phosphatase of interest (Bendall et al., 2011; Budnik et al., 2018; Lun et al., 2017).

Mass cytometry allows simultaneous quantification of over 40 proteins and protein modifications at single-cell resolution and thus enables profiling of complex cellular behaviors in highly heterogeneous samples (Bendall et al., 2011; Bodenmiller et al., 2012; Chevrier et al., 2017b). We have recently established and thoroughly validated an approach that couples transient protein overexpression to mass cytometry-based single-cell analysis and have revealed that protein overexpression induces signaling network modulations in an abundance-dependent manner (Lun et al., 2017).

Here, we applied this technique in a human kinome- and phosphatome-wide screen. To achieve this, we generated a library of DNA vectors encoding 541 GFP-tagged kinases and 108 GFP-tagged phosphatases and transfected these vectors into human embryonic kidney HEK293T cells. Single-cell signaling states were determined by simultaneous measurement of 30 phosphorylation sites known to be involved in regulation of growth, proliferation, survival, and stress signaling pathways. Over 10 million individual cells were analyzed in the 659 overexpression conditions with or without 10-minute EGF stimulation, averaging approximately 7,000 measured cells per sample. Assessing the dependence of each measured phosphorylation site on kinase or phosphatase abundance using BP-$R^2$, a measure to quantify

signaling relationships in single cell analysis (Lun et al., 2017), we identified 1,323 pairs of strong signaling relationships (determined using 108 control experiments). This analysis enabled a functional classification of human kinases and phosphatases based on their abundance-dependent impacts on the signaling network. Furthermore, 208 pairs of previously unknown signaling relationships were identified when compared to the OmniPath database (Türei et al., 2016). By characterizing signaling response dynamics in a follow-up 1-hour EGF stimulation time course, we demonstrated ligand-independent MAPK/ERK activation induced by tyrosine kinase overexpression. We found that in melanoma A375 cells this activation gives rise to $BRAF^{V600E}$ inhibitor resistance. Our screen also revealed that overexpression of ERK-specific phosphatases sustained cell proliferative signals. We confirmed this pro-cancer signaling response using a novel kinase-phosphatase combinatorial overexpression assay.

## Results

### ***Analysis of the human kinome and phosphatome to study the overexpression effects on signaling network states***

Protein abundance variance is often observed in tumors as heterogeneous genomic abnormalities accumulate (Du and Elemento, 2015). We observed up to 50-fold differences in kinase and phosphatase mRNA expression levels among tumor samples in bulk measurements (Figure S1A). This inter-tumoral heterogeneity presumably results in highly variable signaling network states and responses to stimuli or treatments. In addition, a high degree of intra-tumoral expression heterogeneity further challenges cancer therapeutic interventions (McGranahan and Swanton, 2017; Patel et al., 2014). To understand the signaling network modulation in cells that overexpress a defined kinase or phosphatase at various levels, we applied our abundance-dependent signaling network measurement system (Lun et al., 2017) in a kinome- and phosphatome-wide screen.

We cloned open reading frames (ORFs) from the human kinase library (Johannessen et al., 2010) and the human phosphatase library into a vector, enabling expression of GFP-tagged proteins (Couzens et al., 2013). The generated 541 kinase and 108 phosphatase expression clones were transiently transfected into HEK293T cells, individually. Unstimulated cells and cells stimulated for 10 minutes with EGF were harvested and processed with a 126-plex barcoding strategy (adapted from Bodenmiller et al., 2012; Zunder et al., 2015) for simultaneous antibody staining and multiplexed mass cytometry measurement (Figure 1A).

Transient transfection generates a gradient of the GFP-tagged protein of interest (POI) expression levels with a range of up to $10^3$-fold enhancement relative to endogenous POI expression range (Lun et al., 2017). The abundance variation of each overexpressed kinase or phosphatase was quantified by mass cytometry with detection by a metal-conjugated anti-GFP antibody. In addition, we simultaneously quantified 30 phosphorylation states of proteins involved in key cancer-related signaling pathways, including the AKT, PKC, STAT, MAPK/ERK, and stress pathways, and five non-signaling markers to indicate cell physiological states (Table S1). Relationship strength between the abundance of GFP-tagged POI and a measured phosphorylation site was quantified by the binned pseudo-$R^2$ (BP-$R^2$) method (Lun et al., 2017).

We analyzed 108 control samples (FLAG-GFP overexpression or untransfected cells) and used the highest BP-$R^2$ score (0.13) of all controls as the cutoff to consider a signaling relationship as "strong". In total, our human kinome and phosphatome analysis detected 1,323 pairs of strong relationships between POIs and phosphorylation sites (Figure 1B). Among the 649 kinases and phosphatases, 327 had at least one strong signaling relationship (BP-$R^2$ > 0.13) to the cancer-related signaling network when overexpressed. Of these, 245 had narrow influences with modulation of one to five signaling nodes, and 26 overexpressed proteins had broad effects on the network with more than ten of measured phosphorylation sites influenced (Figure S1B). We identified 52 kinases or phosphatases that specifically regulated the AKT pathway when overexpressed. Of the 132 POIs that had abundance-dependent effects on the MAPK/ERK pathway, the majority (104 proteins) also initiated cellular stress responses. We also identified 49 proteins that impacted all of the measured signaling pathways, including 11 receptor proteins (e.g., MET, FGFR1, and PDGFRA), and many MAPK cascade activators (e.g., MAP4K1, MAP4K2, and MAP4K5) (Figure 1C).

### ***Functional classification of kinases and phosphatases based on signaling network modulations***

Our kinome and phosphatome screen characterized effects induced by hundreds of POIs on 30 phosphorylation sites in the cancer-related signaling network with and without EGF stimulation (a total of 60 parameters), yielding an unprecedented view on the effects of each kinase and phosphatase. We indicated the sign for signaling relationships (according to their directionality) to the BP-$R^2$, generating signed-BP-$R^2$ scores that were used for subsequent analyses (Table S2, Methods). To understand the regulatory and functional similarity of overexpressed POIs, we

first applied the dimensional reduction algorithm t-SNE (Maaten and Hinton, 2008) to the matrix of all 60 measured signaling parameters (as signed-BP-$R^2$ scores) over the 327 signaling network-influential kinases and phosphatases (Figure 1D). As expected, homologous groups (paralogs) of kinases and of phosphatases showed nearly identical influences on signaling and overlapped with each other on the t-SNE plot (Figure 1D, green boxes). This demonstrates that our method sensitively, specifically, and reproducibly detected abundance-modulated signaling behaviors. Interestingly, all the eight overexpressed SRC family members, SRC, YES1, BLK, LCK, LYN, HCK, FGR, and FRK, co-localized in the t-SNE analysis (Figure 1D, purple box), indicating that these kinases have similar functions despite the previously revealed differential patterns of expression (Parsons and Parsons, 2004). Members of protein tyrosine phosphatase (PTPN1, PTPN2, and PTPN5) and dual-specificity phosphatase (DUSP3, DUSP4, DUSP6, DUSP7, DUSP10, and DUSP16) families were grouped together, suggesting similarities in regulating the measured cancer signaling network (Figure 1D, orange box).

We then applied hierarchical clustering based on signed-BP-$R^2$ scores of all measured phosphorylation sites (Figure S1C) to analyze functional similarities among all kinases and phosphatases. Based on this analysis we identified 10 major signaling response clusters (color-coded on the t-SNE plot in Figure 1D). Correspondence analysis was performed between these identified clusters and the previously described classes of kinases and phosphatases based on their sequences of the catalytic domain (Johannessen et al., 2010; Sacco et al., 2012a) (Figure S1D). In certain cases, proteins with partial sequence identity had similar influences on signaling. For example, all kinases in cluster 1 were receptor or non-receptor tyrosine kinases (Figure S1D). These kinases were early responders to stimuli, with pleiotropic functions in the signaling networks, as shown in the literature-based graph of canonical EGFR networks (Figure 1E). Clusters 5, 9, and 10 include non-receptor serine/threonine kinases and kinases classified in the group of “other” (i.e., kinases do not fit into any of the major groups) (Figure S1D). Despite conserved catalytic domain sequences, kinases in clusters 5, 9, and 10 induced different cellular responses (Figure 1E). Overexpression of members in cluster 5 activated the PDK1/AKT pathway. Cluster 10 kinases had positive signaling relationships to p90RSK (Ser380), p70S6K (Thr389), PDK1 (Ser241), and MEK1/2 (Ser221) as well as many STAT proteins. Overexpression of cluster 9 components affected proteins involved in the stress response, including MKK4/7 (Ser257/Thr261), p38 (Thr180/Tyr182), and JNK (Thr183/Tyr185). Cluster 9 kinases, when overexpressed, also weakened signaling relationships to MAPK/ERK pathway members, such as MEK1/2 (Ser221), ERK1/2 (Thr202/Tyr204), and p90RSK (Ser380)

after EGF stimulation (Figure 1E). Cluster 7 proteins had negative relationships with the signaling mediators of MAPK/ERK pathway when cells were treated with EGF (Figure 1E). Cluster 7 mostly consists of protein tyrosine phosphatases, but also includes a few proteins from the classes of non-receptor serine/threonine kinase and lipid kinases (Figure S1D). Clusters 2, 3, and 4 consist of kinases and phosphatases from multiple sequence-based classes (Figure S1D). This suggests that these proteins induce similar signaling outcomes despite the differences in catalytic domain sequences. In summary, the human kinome- and phosphatome-wide overexpression analysis identified 10 clusters of kinases and phosphatases that partially matched the sequence-based classification. Distinct signaling patterns were found for each cluster.

***Functional enrichment analysis for identified kinase and phosphates clusters***

Our analysis indicated that kinases and phosphatases with different catalytic domain sequences could impact signaling networks similarly when overexpressed. To understand this, we performed functional enrichment analysis using the STRING database (Szklarczyk et al., 2017) on the 10 identified clusters (Figure 1D and 2A). We found that seven of the 10 clusters had significant functional enrichment ($p < 0.01$, Table S3, statistical details in Methods). For each of these clusters, the top five specific functions are shown (Figure 2A). Physical and functional interaction enrichments are shown as protein-protein association networks for clusters 1 and 7 as examples in Figures 2B and 2C and for all other clusters in Figure S2.

As expected, proteins in cluster 1 are enriched for membrane and non-membrane tyrosine kinases with autophosphorylation ability (Figure 2B) with the significance being robust after the removal of redundancy (homology) effects (Table S3, Methods). Although these kinases are closely associated (Figure 2B), we did not find significantly enriched terms for specific signaling pathways or physiological functions (Table S3). Given that all tyrosine kinases of this group generate highly similar signaling events in the measured network (Figure S1C), other factors such as their expression patterns or regulatory protein complexes likely drive their diverse functions in regulating cell phenotypes.

Cluster 7 is enriched for protein tyrosine phosphatases that negatively regulate MAPK pathways. Intriguingly, several kinases are present in this cluster as well, including KSR1 and ARAF, which have similar negative regulatory effects on the MAPK/ERK signaling pathway (Figure 2C). KSR1 and ARAF are core components of the KSR-RAF dimeric protein complex that

transduces signal in the MAPK/ERK cascade (Lavoie and Therrien, 2015). Overexpressing one subunit of this protein complex might result in competitive inhibition, diminishing the downstream signal activities in a similar manner as phosphatase overexpression. These analyses demonstrated that proteins with different catalytic functions can mediate highly related signaling responses when overexpressed.

Other clusters were enriched for homologous kinases (clusters 5 and 10) and the protein phosphatase type 2A complex (cluster 2) (Figure S2). We also found that certain kinases associated with the same signaling pathway modulated the measured network differently when overexpressed (Figure S2). Examples are MAPK1 (cluster 6) and RPS6KA1 (cluster 10), which regulate the MAPK/ERK pathway, and PDPK1 (cluster 10) and AKT1 (cluster 4), which regulate the PDK1/AKT pathway. These results suggest that kinase overexpression impacts signaling networks differently to direct kinase activation. Thus, variances in overexpression-induced signaling network modulation are not fully explained by the catalytic functions of kinases and phosphatases. In cells, signaling pathway activity is determined not only by phosphorylation but also by many intrinsic factors, such as protein subcellular localization and protein complex formation.

***Detecting novel signaling relationships from the kinome- and phosphatome-wide analysis***

The functions of many kinases and phosphatases analyzed in our screen are unknown or only poorly characterized. We therefore hypothesized that our global analysis could lead to the identification of novel signaling relationships. To assess this, we performed a systematic comparison between all identified overexpression-induced signaling relationships and records in OmniPath, an integrated database of literature-curated signaling interaction information (Türei et al., 2016). We first mapped all pairs of relationships to the OmniPath signaling network, then computed the signed, directed paths for each pair of relationship (Krumsiek et al., 2011; Perfetto et al., 2016). The distance between an overexpressed protein and a measured phosphorylation site is represented by the length of the path (Figure 3A). For example, a distance of zero indicates the relationship between the overexpressed POI and its own phosphorylation levels. Of 14 pairs of signaling relationships with a known distance of zero, with and without 10-minute EGF stimulation, 12 had strong BP-$R^2$ values (Figure 3A), revealing that the phosphorylation abundance of kinases is often determined by their own expression level, even in the absence of additional perturbation.

We detected 208 (16%) strong relationships (BP-$R^2$ > 0.13) with infinite distance (Figure 3A, Table S4), indicative of connections not described previously. In total, 93 overexpressed POIs contributed to these signaling relationships, which were enriched (in absolute count) in clusters 2, 3, and 4, and to a less extend in cluster 6 (Figure 3B). We did not detect any relationship with infinite distance in clusters 9 and 10 (Figure 3B); POIs from these clusters participate in MAPK signal transduction (Figure 2A), which is one of the best studied signaling pathways. We also assessed the distribution of infinite paths for each sequence-based kinase and phosphatase class and did not detect any enrichment (Figure S3A). There were also 132 pairs of strong relationships between proteins with length of signed directed path above three in OmniPath, suggesting potentially undiscovered direct or short-range connections (Figure 3A).

Many of the identified novel signaling relationships (i.e., those with infinite path length) and the associated POIs were related to disease and to newly identified kinases (Figures 3C, 3D, S3B, and S3C). For instance, RIOK2 (highlighted in Figure 3C) has been recently shown to correlate with poor prognosis of patients with non-small cell lung cancer, but the underlying signaling mechanisms are unclear (Liu et al., 2016). We discovered that RIOK2 overexpression impacted several phosphorylation sites, most strongly Thr172 on AMPK$\alpha$, Ser257/Thr261 on MKK4/7, and Thr180/Tyr182 on p38 (Figure 3C), indicating the activation of the AMPK/p38 axis upon RIOK2 overexpression. The AMPK/p38 axis regulates cellular energy metastasis, contributing to cancer cell survival in nutrient-deficient conditions (Chaube et al., 2015; Zadra et al., 2015). We also found that MGC42105 (NIM1K) (highlighted in Figure 3D), a newly identified kinase in cluster 5, regulated the AKT pathway. Overexpression of MGC42105 altered phosphorylation of Ser241 on PDK1, Thr389 on p70S6K, and Ser235/Ser236 on S6. Overexpression of MGC42105 also contributed to the activation of stress pathways, as strong relationships to p-p53 (Ser15) and p-AMPK$\alpha$ (Thr172) were observed (Figure 3D). In summary, mapping our identified signaling relationships to the OmniPath database enabled the identification and assignment of novel signaling functions to many kinases and phosphatases. We also reveal potential novel signaling mechanisms associated with poor prognosis of cancer patients (e.g., for RIOK2).

***In-depth analysis of signaling dynamics reveals overexpression-dependent MAPK/ERK activity***

Signaling dynamics are essential for understanding of diseased signaling circuits within a network and in the prediction of drug efficiency (Du and Elemento, 2015; Hughey et al., 2009). We have previously shown that altering expression levels of central signaling proteins in the EGFR signaling network results in complex changes in network dynamics (Lun et al., 2017). Given the key role of signaling dynamics on cell proliferation, growth, and differentiation (Koseska and Bastiaens, 2017), we systematically evaluated kinases and phosphatases from the 10 identified clusters for overexpression-induced signaling dynamic modulations. We calculated the differences in signed-BP-$R^2$ scores between the EGF-stimulated (10 min) and unstimulated conditions to identify cases in which overexpression modulates signaling dynamics (i.e., the strength and the shape of abundance-dependent signaling relationship changes between the unstimulated and the 10-minute EGF stimulated conditions). We found that POIs in clusters of 1, 6, 7, 9, and 10 strongly modulated signaling network dynamics when overexpressed (Figure S4A). We then analyzed the overexpression effects of the 39 strongest signaling dynamics influential POIs (criteria described in Methods) over a 1-hour EGF stimulation time course in depth (Figures 4 and S4). The dynamic responses of all measured phosphorylation sites are shown in Figure 4A. Features of the signaling dynamics, such as signaling amplitudes and peak times (see Methods) are shown in Figures S5 and S6.

Hierarchical clustering of the overexpression-induced EGFR signaling dynamics of the 39 selected proteins revealed six groups (Figure 4A). Each of these six groups typically reflected one or two of the 10 major signaling response clusters identified in the previous sections (Figure 1D); the correspondence is shown in Figure S4B. Signaling network responses for one representative kinase or phosphatase from each of the six identified groups are illustrated in Figure 4B. Interestingly, the EGF stimulation-influenced relationships were mostly observed for phosphorylation sites within the MAPK/ERK signaling cascade, including p-MEK1/2 (Ser221), p-ERK1/2 (Thr202/Tyr204), p-p90RSK (Ser380), and p-S6 (Ser235/Ser236), rather than AKT, PKC, STAT, or stress response pathways (Figure 4A). Given that the MAPK/ERK is the major proliferative pathway that is known to be involved in tumor progression and drug response, we focused our subsequent analyses on this pathway to identify novel regulatory mechanisms potentially relevant to cancer.

EGFR overexpression (as an example of group B) mediated activation of the MAPK/ERK pathway (Figure 4B), multiple STAT kinases, PLCγ2, and the stress signaling pathways. Phosphorylation of Thr202/Tyr204 on ERK1/2 was elevated in cells with high levels of GFP-tagged EGFR in the absence of EGF stimulation (Figure 4C). These cells responded weakly to EGF stimulation, indicating ligand-independent ERK activation (Figures 4C and S5B). Overexpression of other members of group B, including MST1R, MET, FGFR1, TYRO3, TEC, and ABL1, influenced signaling dynamics of p-ERK1/2 and p-90RSK in a manner similar to that of EGFR overexpression (Figure 4A). These proteins correspond to the global cluster 1 (Figure S4B), and members of this class mediate oncogenic signaling for many cancer types (Duan et al., 2016; Paul and Mukhopadhyay, 2004; Salgia, 2017). Taken together, this suggests that ligand-independent MAPK/ERK signaling activation causes uncontrolled cell proliferation.

Interestingly, in the absence of EGF stimulation, overexpression of phosphatases DUSP4 (from group E) and PTPN2 (from group F) did not affect levels of phosphorylation in the MAPK/ERK pathway (Ser221 on MEK1/2, Thr202/Tyr204 on ERK1/2, or Ser380 on p90RSK) (Figures 4B, 4D-4G, S4C, and S4D). This suggests either a mechanism compensates for phosphatase overexpression to maintain basal MAPK/ERK signaling or the overexpressed phosphatases are inactive without EGF stimulation. Upon EGF stimulation, signaling dynamics on phosphorylation sites of the MAPK/ERK pathway were modulated by DUSP4 and PTPN2 in an abundance-dependent manner, as negative signaling relationships to p-ERK1/2 and p-p90RSK were detected (Figures 4B, 4D, 4E, S4C and S4D). DUSP4 or PTPN2 overexpression also resulted in reduced signaling amplitudes on p-ERK1/2 and p-p90RSK (Figure S5C). Intriguingly, different p-MEK1/2 dynamics were observed in cells overexpressing these two phosphatases. There was a strong positive relationship between GFP-tagged DUSP4 and p-MEK1/2 beginning at 10 minutes after EGF addition (signed-BP-$R^2$ = 0.23) with relationship strength constantly increasing until the 30-minute time point (Figure 4F), indicating a more sustained MEK1/2 phosphorylation in cells with higher levels of DUSP4 than in cells with lower DUSP4 expression. In contrast, a negative relationship between GFP-tagged PTPN2 and p-MEK1/2 was observed during EGF stimulation with a strong relationship at the 5-minute (signed-BP-$R^2$ = -0.23) and 10-minute (signed-BP-$R^2$ = -0.14) time points (Figure 4G). DUSP4 is known to selectively de-phosphorylate ERK1 and ERK2 (Guan and Butch, 1995). Our data indicate that during EGF stimulation, DUSP4 overexpression diminishes ERK1/2 phosphorylation, and, in turn, the negative feedback from ERK1/2 to MEK1/2 is likely attenuated, resulting in constant activation of MEK1/2. Substrates of PTPN2 are primarily membrane kinases including EGFR (Mattila et al.,

2005). Overexpression of PTPN2, therefore, downregulates activation states of all measured signaling proteins known to be downstream of EGFR, including MEK1/2 and ERK1/2 (Figure 4h). Systematic analysis of all overexpressed phosphatases over the 1-hour time course after EGF addition confirmed that all other phosphatases in group E (DUSP6, DUSP7, DUSP10, DUSP16, PTPN5, and PTPN7) target phosphorylation sites of Thr202/Tyr204 on ERK1/2, thereby decreasing the negative feedback from ERK1/2 to MEK1/2 and causing sustained MEK1/2 activation (Figures 4A and 4H).

In conclusion, these signaling dynamic analyses indicate that first, overexpression of tyrosine kinases including EGFR induce EGF ligand-independent MAPK/ERK signaling activation. Second, phosphatases differentially regulate MAPK/ERK signaling responses determined by their substrate specificities. Third, the ERK-specific phosphatases control the strength of the negative feedback loop from ERK1/2 to MEK1/2 in an abundance-dependent manner.

### ***Pairwise overexpression analysis reveals phosphatase sustains the kinase-induced MAPK/ERK signaling***

Phosphatase overexpression is oncogenic in different tumor types, but the signaling mechanisms remain unclear (Julien et al., 2007, 2011). Recent work indicates that overexpressed phosphatases increase the malignancy of cancers that have a hyper-activated MAPK/ERK pathway (Julien et al., 2007; Low and Zhang, 2016; De Vriendt et al., 2013). Our result in the previous section suggested a mechanism through which overexpression of ERK-specific phosphatases sustains MEK phosphorylation levels (Figures 4F and 4H). To assess whether an additional, secondary signaling input that increases MAPK pathway activity could lead to oncogenic-like signaling, we developed a combinatorial transfection assay in which overexpression of a kinase and a phosphatase were detected via a FLAG-tag and a GFP-tag, respectively, providing a two-dimensional analysis of abundance-dependent signaling modulation on the single cell level (Figure 5A). Using this approach, we analyzed the MAP2K2, MAPK1, and RPS6KA1 (also known as MEK2, ERK2, and p90RSK1) kinases, and the DUSP4, DUSP7, and PTPN2 phosphatases in nine combinations of double overexpression over a 1-hour EGF stimulation time course (Figure 5B). After measurement, we subdivided cells according to the expression levels of the FLAG-tagged kinase and GFP-tagged phosphatase into 25 bins within the two-dimensional overexpression space. For each bin, signaling states as defined by phosphorylation levels, and signaling trajectories with respect to all individual bins over the 1-hour EGF time course were analyzed.

When overexpressed individually, we observed that DUSP4 overexpression sustained the phosphorylation of Ser221 on MEK1/2 over the 1-hour EGF stimulation time course likely due to the weakened ERK to MEK negative feedback (Figures 5C and 5D). MAP2K2 (MEK2)-FLAG overexpression led to an increased MEK1/2 phosphorylation (Figure 5C). Interestingly, combined signaling inputs from MAP2K2-FLAG and DUSP4-GFP co-overexpression further increased the hyper-activated states of MEK1/2 over the 1-hour EGF stimulation time course compared to the activation induced by MAP2K2-FLAG overexpression alone (Figures 5C-5E, brown arrow). Moreover, in cells with simultaneously overexpressed MAP2K2-FLAG and DUSP4-GFP, the downstream ERK1/2 phosphorylation on Thr202 and Tyr204 were inhibited (Figures 5C-5E). Highly activated MEK1/2 could lead to ERK-independent oncogenic-like signaling as revealed previously (Burgermeister and Seger, 2008; Takahashi-Yanaga et al., 2004). In contrast, double overexpression of MAP2K2-FLAG and PTPN2-GFP (the latter a phosphatase targeting EGFR) did not sustain the MEK1/2 signaling; rather, it dampened the MEK1/2 signaling amplitudes in response to EGF stimulation (Figure S7).

Overexpression of FLAG-tagged MAPK1 (ERK2) drastically augmented ERK1/2 phosphorylation state during EGF stimulation (Figure 5F), increased p-ERK1/2 amplitudes, and delayed p-ERK1/2 peak times (Figure 5G). These results are in agreement with findings of previous studied MAPK1 overexpression effects (Lun et al., 2017). The simultaneous overexpression of MAPK1-FLAG and DUSP7-GFP decreased p-ERK1/2 levels at all time points and reduced the signaling amplitudes. Further, DUSP7 delayed p-ERK1/2 peak times upon EGF stimulation: in cells with the highest MAPK1 abundance and mid-level overexpression of DUSP7, ERK1/2 phosphorylation peaked at 30 minutes after EGF addition (Figures 5F and 5G, purple arrow), whereas in untransfected cells, p-ERK1/2 peaked at the 5-minute time point (Figures 5F and 5G). As expected, DUSP7 overexpression also resulted in constant MEK1/2 phosphorylation (Figures 5F and 5G, green arrow). Compared to cells only overexpressing MAPK1 (ERK2), which induced strong but transient ERK activation, the additional low-to-mid levels of DUSP7 decreased the ERK1/2 phosphorylation amplitude and partially limited the negative feedback signal from ERK to MEK, inducing a sustained MEK activation and prolonged ERK signal.

Assessing signaling responses in the pairwise overexpression assay, we demonstrated kinase and phosphatase coregulatory mechanisms in the MAPK/ERK cascade. We confirmed that

phosphatase overexpression does not dampen signaling activity at the steady states, even with activated oncogenic signaling. Further, our analysis indicates that the overexpression of certain phosphatases, such as DUSP4 and DUSP7, lead to sustained activation of ERK due to the reduced negative feedback strength. This mechanism might underlie the pro-cancer effects of phosphatase overexpression.

***Signaling relationship to p-ERK1/2 predicts overexpression-induced vemurafenib resistance in melanoma A375 cells***

As protein overexpression has been correlated with drug resistance of cancer cells (Johannessen et al., 2010; Shaffer et al., 2017), we next sought to determine whether our kinome- and phosphatome-wide signaling network profiles could identify kinases or phosphatases that, when overexpressed, induce drug resistance; these enzymes can be potential biomarkers of drug resistance. In melanoma cells carrying the $BRAF^{V600E}$ mutation, overexpression of certain kinases is associated with *de novo* or acquired resistance to RAF inhibition; Johannessen et al. has identified nine kinases that drive resistance when overexpressed in a cell-based assay (Johannessen et al., 2010). Eight of these proteins were analyzed in our screen, and interestingly, six had abundance-dependent signaling modulations to p-ERK1/2 (Thr202/Tyr204) in unstimulated cells (Figure S8A). The 10-minute EGF stimulation reduced relationship strengths for each of these six kinases (Figure S8B), indicating that these overexpression-related ERK activations were ligand-binding independent (i.e., cells with high POI-GFP levels did not respond to EGF stimulation, Figure 4C). These data suggest that overexpression of kinases that induce ligand-independent MAPK/ERK pathway activation might underlie resistance to $BRAF^{V600E}$-targeted inhibitors.

To determine whether signaling relationships to p-ERK1/2 in our kinome- and phosphatome-wide analysis have potential as biomarkers for drug resistance, we examined kinases in our screen with the highest signed-BP-$R^2$ values in relation to p-ERK1/2 in the absence of EGF stimulation, including ABL1, BLK, FES, MAP3K2, MAP3K8, MOS, NTRK2, SRC, and YES1. These kinases, and $MEK1^{DD}$ as positive control, were overexpressed in A375 cells that express $BRAF^{V600E}$. Cells were subsequently treated for 48 hours with the $BRAF^{V600E}$ inhibitor vemurafenib or DMSO as a control (Figure 6A, all experiments performed in three replicates). In A375 cells, the BRAF substrate MEK1/2 is activated and basal levels of p-ERK1/2 are high. As expected, overexpression of $MEK1^{DD}$, which constitutively phosphorylates Thr202 and Tyr204 on ERK1/2 (Johannessen et al., 2010; Lun et al., 2017), only slightly enhanced ERK1/2

phosphorylation in cells treated with DMSO (Figure 6B). A weak signaling relationship was observed between GFP-tagged $MEK1^{DD}$ and p-ERK1/2 with a signed-BP-$R^2$ of 0.11 (Figure 6B, left). Compared to the DMSO control, treatment with vemurafenib inhibited $BRAF^{V600E}$ activity and reduced p-ERK1/2 levels in untransfected cells and in cells with low $MEK1^{DD}$ expression levels. However, cells with high levels of $MEK1^{DD}$ were insensitive to vemurafenib, such that hyper-phosphorylated ERK1/2 remained after the 48-hour treatment (signed-BP-$R^2$ value of 0.53, Figure 6B, right). In the control cells without $MEK^{DD}$ overexpression, the effect of vemurafenib on p-ERK1/2 levels did not alter signaling relationships as quantified by signed-BP-$R^2$ (Figures S8C and S8D).

Similar as observed in cells overexpressing GFP-tagged $MEK1^{DD}$, vemurafenib treatment also increased signaling relationships between all our analyzed GFP-tagged kinases (ABL1, BLK, FES, MAP3K2, MAP3K8, MOS, NTRK2, SRC, and YES1) and p-ERK1/2 (Figure 6C). A closer inspection of the signed-BP-$R^2$ signaling strength of cells treated for 48 hours with vemurafenib and DMSO-treated cells showed that eight of the nine tested kinases had higher values than those of negative controls (Figure 6D). These results indicate that overexpression of ERK-activating kinases limits the effects of vemurafenib in reducing p-ERK1/2 signaling, similarly as the $MEK^{DD}$ positive control.

To assess the survival of cells with different expression levels of each kinase, we assigned each single cell into one of four bins based on the abundance of GFP-tagged POI and calculated the percentage of the number of cells in each bin relative to the total cell count (Figure 6B). As expected, the positive control cells that expressed $MEK1^{DD}$-GFP had significant cell enrichment in the fourth bin (i.e., the bin with the highest expression level of the GFP-tagged POI), with four times higher percentage of cells at this level after 48 hours of vemurafenib treatment compared to the DMSO-treated control (Figure 6E). Similarly, all nine examined kinases showed enrichment of cell abundance in the fourth bin: in six cases, this enrichment was statistically significant (Figure 6E).

These results indicate that, in melanoma A375 cells, overexpression of kinases capable of ligand-independent ERK activation reduces cellular dependency on signaling inputs from $BRAF^{V600E}$ for proliferation, rendering the cells less sensitive to the $BRAF^{V600E}$ inhibitor vemurafenib (Figure 6F). In the absence of inhibitor, however, cells overexpressing ERK-activating POIs do not acquire proliferative advantage (Figure 6F). As has been shown

previously, the hyperactivity of ERK induces cell senescence and apoptosis (Cagnol and Chambard, 2010; Xu et al., 2014). Our kinome and phosphatome screen detected 54 POIs that, when overexpressed, caused EGF-independent ERK activation (Table S5). These proteins can potentially be used as biomarkers to predict resistance to $BRAF^{V600E}$ inhibitors in melanoma patients carrying BRAF mutations.

**Discussion**

The data described here are unique for the broad coverage of the human kinome and phosphatome, the multiplexed measurement of cellular phosphorylation states and dynamics at single-cell resolution, and the wide abundance range (over three orders of magnitude) over which proteins of interest were studied. Our analyses enabled protein abundance-determined functional classification, signaling kinetics quantification, and the identification of potential biomarkers of drug resistance.

Extending our previously established approach (Lun et al., 2017) for analysis of the dependence of signaling behaviors on protein expression levels to the human kinome and phosphatome, we transiently transfected a library of 649 plasmids encoding GFP-tagged kinases and phosphatases individually into HEK293T cells, yielding gradient overexpression levels for each GFP-tagged POI. The abundance of each overexpressed kinase or phosphatase was quantified simultaneously with 30 cancer-related phosphorylation sites and five non-signaling markers (Table S1) using a mass cytometry-based multiplexed single-cell assay. Signaling states and dynamics over the expression continuum for every analyzed POI were comprehensively profiled. We applied the recently developed statistical measure, BP-$R^2$, to quantify signaling strength between each GFP-tagged POI and each measured phosphorylation site (Lun et al., 2017) and classified all overexpressed kinases and phosphatases based on their abundance-determined signaling network states with or without 10-minute EGF stimulation.

Protein abundance and mRNA expression levels of kinases and phosphatases have been quantified in normal and diseased tissues by multiple approaches (Petryszak et al., 2016; Uhlen et al., 2017; Wang et al., 2015). Our analysis, for the first time, characterized at kinome- and phosphatome-wide scope how these proteins differentially modulate network behaviors when expressed over a concentration gradient. In the overexpression effect-based functional classification, we assigned kinases and phosphatases into 10 clusters, each with a distinct

function in signaling transduction. These clusters only partly agreed with the kinase and phosphatase classes based on their sequence of catalytic domain, indicating the dissimilar network alterations between signaling protein overexpression and activation. Sequence-based classification considers the catalytic specificities of kinases and phosphatases. However, by altering the signaling protein expression levels (or concentration), dynamics of both upstream and downstream signaling events, and the assembly of multiprotein complexes can be modulated, resulting in more complicated network changes.

Interestingly, our kinase and phosphatase classification and the functional enrichment analysis suggest that kinases, such as KRS1 (a pseudo-kinase) and ARAF, negatively regulate MAPK/ERK signaling when overexpressed, similarly to many tyrosine phosphatases (Figures 1D, 2A, and 2C). In the MAPK/ERK cascade, dimerization between KSR proteins (KSR1 and KSR2) and RAF proteins (ARAF, BRAF, and CRAF), is required for the activation of the downstream kinase MEK (Lavoie and Therrien, 2015; Rajakulendran et al., 2009). Overexpressing either of KSR1 or ARAF leads to competitive protein binding that disrupts the formation of the dimeric protein complex. Interestingly, cells with CRAF (also known as RAF1) or KSR2 (both in cluster 6) overexpression showed less degree of MAPK/ERK signaling attenuation (Figure S1C), compared to those with KSR1 or ARAF overexpression. This suggests the differential capability of individual RAF proteins or KSR proteins in forming homodimers that can compensate overexpression-induced signaling disruption.

Phosphatase overexpression is observed to drive tumor progression with unclear signaling mechanisms (Julien et al., 2011; Low and Zhang, 2016; De Vriendt et al., 2013). We indicated that rather than directly activating a cancer-driven signaling pathway, overexpression of ERK-specific phosphatases modulates signaling dynamics that leads to prolonged proliferative signal in cells. Phosphatases have been suggested in many recent studies as potential therapeutic targets (Bollu et al., 2017; Julien et al., 2011; Low and Zhang, 2016). Our result indicates again the importance of developing phosphatase inhibitors for cancer treatment.

Following a pilot kinome study on overexpression-related resistance to RAF inhibition (Johannessen et al., 2010), we discovered that ligand-independent ERK activation induced by kinase overexpression to be the underlying signaling mechanism leading to above drug resistance. Our kinome and phosphatome analysis further identified 54 proteins that caused ligand-independent ERK activation when overexpressed. These proteins were then suggested

as potential biomarkers for RAF inhibitor resistance. Mass cytometry-based single cell analysis enables identifying differed signaling responses to a drug treatment in rare cell populations. It is therefore more sensitive in identifying biomarkers for overexpression-induced drug resistance, compared to the previous population-based assay (Johannessen et al., 2010). Our data can be further used to reveal signaling mechanisms in diseased cellular behaviors that are caused by kinase or phosphatase overexpression. Then, following the discovered pathological signaling behaviors, more biomarkers can be identified for the given disease, based on our kinome- and phosphatome-wide analysis.

Our analysis has several limitations. First, the measured overexpression effects can be indirect (i.e., a protein overexpression might lead to cellular stress that activates MAPK/p38 and MAPK/JNK cascades). However, indirect effects are also important signaling responses that can be indicative for any cells with such overexpression. Second, our mass cytometry-based analysis applies pre-selected antibodies targeting 30 specific phosphorylation sites that does not capture all signaling responses induced by overexpression of a kinase or a phosphatase. Nevertheless, our antibody panel is designed based on literature information to cover the most critical and informative phosphorylation sites in the cancer-related signaling network. Third, the GFP-tag can potentially disrupt the function of a kinase or phosphatase. Likely, a non-functional protein does not induce specific signaling network modulation and should therefore yield weak BP-$R^2$ values ($\leq$ 0.13) only.

In summary, we demonstrated, in the human kinome- and phosphatome-scale analysis, how overexpression of each signaling protein modulates signaling networks in an abundance-dependent manner. This establishes that protein expression levels can result in different signaling states in a heterogeneous population. Our analysis expands the functional classification of the human kinases and phosphatases, and suggests 208 novel signaling relationships that can be interrogated to improve our understanding of signaling causality and network structure. We showed novel oncogenic-like signaling mechanisms and identified cancer biomarkers with our analysis. Our data are also suitable for the inference of signaling pathway kinetics using mathematical models and for the development of network reconstruction methods.

## Author contributions

X.-K.L. and B.B. conceived the study. X.-K.L. performed experiments, data processing, and data analysis. D.S. performed functional enrichment and functional association analysis. A.G. performed novel relationship analysis. N.D. cloned the phosphatase expression library. V.Z. implemented the BP-$R^2$ platform and helped with data analysis. X.-K.L., D.S., A.G., N.D., V.Z., J.S.-R., C.M., and B.B. performed the biological interpretation. X.-K.L. and B.B. wrote the manuscript with input from all authors.


## Acknowledgement

We would like to thank the Bodenmiller lab for support and fruitful discussions, the Sommer laboratory for sharing experimental materials, the Lehner lab and the Mosimann lab for sharing equipment, Dr. Vinko Tosevski and Dr. Tess Brodie at the Mass Cytometry Facility, University of Zürich for support and troubleshooting help. We would especially like to thank Dr. A.-C. Gingras, Lunenfeld-Tanenbaum Research Institute, for sharing the pDEST vectors used in this study. This work was supported by an SNSF R`Equip grant, a SNSF Assistant Professorship grant (PP00P3-144874), by the European Research Council (ERC) under the European Union's Seventh Framework Program (FP/2007-2013)/ERC Grant Agreement n. 336921 and an NIH grant (UC4 DK108132).


## Figure Legend

***Figure 1.*** **Kinome and phosphatome-wide screen for effects of protein abundance on signaling states and dynamics.**

**A,** The experimental workflow: ORFs of 541 human kinases and 108 human phosphatases were cloned into a vector to enable expression of GFP-tagged fusion proteins upon transient transfection into HEK293T cells. Cells with or without 10-minute EGF stimulation were harvested, barcoded, and stained with antibody mix before mass cytometry-based single cell analysis. **B,** Plot of counts vs. BP-$R^2$ values for control and experimental samples. Cut-off value was determined by analysis of the BP-$R^2$ values in all control samples. **C**, Venn diagram showing the quantification of POIs with abundance-dependent influences on the AKT pathway (p-PDK1, p-GSK3$\beta$, $\beta$-catenin, p-mTOR, p-p70S6K, p-4EBP1, and p-S6), MAPK/ERK pathways (p-RAF, p-MEK1/2, p-ERK1/2, p-p90RSK, p-CREB, and p-SMAD2/3), stress response pathways (p-MKK3/6, p-MKK4/7, p-p38, p-JNK, p-MAPKAPK2, p-AMPK$\alpha$, and p-p53), PKC

pathway and STAT pathways (grouped for illustration purposes; p-SRC, p-FAK, p-BTK, p-PLCγ2, p-MARCKS, p-NFκB, p-STAT1, p-STAT3, and p-STAT5). **D**, t-SNE analysis of overexpressed kinases and phosphatases performed on signed-BP-$R^2$ of all measured phosphorylation sites with and without EGF stimulation, color-coded by hierarchical clusters. **E,** The mean signed-BP-$R^2$ values of all measured phosphorylation sites in each cluster of kinases or phosphatases shown in a literature-guided canonical signaling network visualization.

***Figure 2.*** **Functional enrichment analysis of kinases and/or phosphatases in each cluster. A,** Unrooted tree shows the hierarchical clustering of the kinases and phosphatases based on their signed-BP-$R^2$ scores. Terms of enriched functions ($p<0.05$) from each cluster are annotated with circle color indicating the p-value and circle size showing the coverage of cluster components. The percentage of associated proteins is indicated by the size of the adjacent circle. **B-C,** Functional association analysis performed with the STRING database (Szklarczyk et al., 2017) for cluster 1 and cluster 7, respectively. Edges with confidence above 0.2 are shown in the network. Functional enrichments are shown as color-coded-pies with the pie radius indicating the p-value.

***Figure 3.*** **Prediction of novel signaling connections by comparison with literature evidence in the signaling interaction database OmniPath** (Türei et al., 2016). **A,** Abundance-dependent relationship strength for each pair of overexpressed POI and measured phosphorylation site, as quantified with signed-BP-$R^2$, plotted on the length of shortest signed directed path between the two extracted from the OmniPath database. **B,** Occurrences of strong signaling relationships (BP-$R^2$ > 0.13) with 0-5 or infinite path length (OmniPath) in each individual hierarchical cluster. **C-D,** For clusters 8 and 5, respectively, shortest signed directed path length for each determined strong signaling relationships shown in Circos plots (Krzywinski et al., 2009).

***Figure 4.*** **Effects of EGF stimulation on 39 kinases and phosphatases. A,** Heat map of signed-BP-$R^2$ scores for measured signaling relationships over a 1-hour EGF stimulation time course. Six identified groups of kinases and phosphatases are labeled in color codes. **B,** For one representative POI from each group, signaling relationships to all measured phosphorylation sites, as quantified by signed-BP-$R^2$, are shown in the literature-guided canonical signaling network map. **C-G,** Violin plots show cell distribution in each of the ten bins divided over the gradients of GFP-tagged POI expression level for EGFR-GFP to p-ERK1/2,

DUSP4-GFP to p-ERK1/2, PTPN2-GFP to p-ERK1/2, DUSP4-GFP to p-MEK1/2, and PTPN2-GFP to p-MEK1/2 over the 1-hour EGF stimulation time course. Medians of all 10 bins are connected to indicate the shape of signaling relationships (black lines) with the relationship strength quantified by signed-BP-$R^2$, as shown on top of each individual plot. **H,** Schematic illustration of how two sets of phosphatases induce different abundance-dependent influences on the signaling dynamics of the MAPK/ERK cascade.

***Figure 5.*** **Pairwise overexpression of two-dimensional protein abundance dependency analysis of signaling behaviors. A,** Workflow of the pairwise overexpression. Two plasmids encoding a FLAG-tagged kinase and a GFP-tagged phosphatase were transfected into HEK293T cells successively. Cells were binned into 25 groups according to their FLAG and GFP abundances. The median level of each measured phosphorylation site was computed for each bin. **B**, Kinases MAP2K2, MAPK1, and RPS6KA1 and phosphatases DUSP4, DUSP7, and PTPN2 were selected for the pairwise overexpression, generating nine overexpression combinations in total. **C,** In cells with pairwise overexpression of a FLAG-tagged kinase MAP2K2 (MEK2) and a GFP-tagged phosphatase DUSP4, median phosphorylation levels of p-MEK1/2 and p-ERK1/2 are plotted for all bins over the 1-hour EGF stimulation time course. **D,** Signaling trajectories of p-MEK1/2 and p-ERK1/2 plotted as the medians of each individual bin over the 1-hour EGF stimulation time course. **E,** Schematic illustrating the modulation of RAF/MEK/ERK cascade signaling states and dynamics by the pairwise overexpression. **F-G,** Analysis of dependence of p-ERK1/2 and p-MEK1/2 signaling on MAPK1(ERK2)-FLAG and DUSP7-GFP.

***Figure 6.*** **Kinase overexpression induces vemurafenib resistance in the melanoma A375 cells by activating ligand-binding independently of ERK signaling. A,** Selected kinases were transfected into A375 cells and cultured in the inhibitor- or DMSO-containing media for 48 hours before the signaling states and enrichment of kinase overexpression cells were assessed. **B,** Data from each sample were divided into four bins depending on the expression level of the GFP-tagged kinase. Signed-BP-$R^2$ analysis was performed to quantify signaling relationships with and without vemurafenib. **C,** Histogram of abundance-dependent relationship strength for each overexpressed POI to p-ERK1/2, quantified as signed-BP-$R^2$ from three replicate experiments. **D,** The mean differences of the three replicates between vemurafenib-treated cells and DMSO-treated cells in their signed-BP-$R^2$ scores with p-ERK1/2. **E,** Proportion of cells in bin 4, cells with the highest levels of POI, for each individual overexpressed kinase in

vemurafenib-treated cells and DMSO-treated cells. ($^{*}$ $p < 0.05$; $^{**}$ $p < 0.01$; $^{***}$ $p < 0.001$, $n = 3$) **F,** Illustrations of how ERK-activating POIs maintain the proliferative state in cells treated with vemurafenib and, in the un-drugged condition, how (despite overexpression of these POIs) over-activated ERK signaling leads to cell senescence and apoptosis.

## Methods

### ***Data availability***

All raw data are available at http://www.cytobank.org/bodenmillerlab.

### ***Cloning***

The human kinase library plasmid kit, containing open reading frames (ORFs) in Gateway Entry vectors, was provided by William Hahn and David Root (Johannessen et al., 2010; Yang et al., 2011) via Addgene (Kit # 1000000014). The human phosphatase library was obtained from Dharmacon (OHS4941, ORFeome Human Entry Collection Phosphatase). Destination vectors, including pDEST pcDNA5 FRT TO-eGFP and pDEST 3' Triple Flag pcDNA5 FRT TO, were kindly provided by Dr. Anne-Claude Gingras at Mount Sinai Hospital, Toronto, Canada (Couzens et al., 2013). Expression vectors encoding the FLAG- or GFP-tagged fusion proteins were generated via Gateway Cloning and sequenced before transfection.

### ***Cell culture***

HEK293T cells, obtained from ATCC, were cultured in DMEM (D5671, SIGMA), supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin. A375 cells, a gift from Dr. Lukas Sommer at University of Zürich, were cultured in RPMI Medium 1640 (21875-034, Gibco) with 10% FBS, 100 U/ml penicillin, and 100 µg/ml streptomycin. For cell passaging or harvesting, HEK293T cells and A375 cells were incubated with 1X TrypLE™ Express (Life Technologies) at 37 °C for 2 minutes and 5 minutes, respectively.

### ***HEK293T cell transfection and stimulation***

HEK293T cells were seeded at the density of 0.7 million per well in 6-well plates. After 24 hours, cells were transfected with 2 µg plasmid and 4 µl of jetPRIME (PolyPlus) per well with the standard protocol provided by the manufacturer. For kinase and phosphatase double transfection experiments, cells were transfected with plasmids encoding GFP-tagged kinases and FLAG-tagged phosphatases 16 hours and 24 hours after seeding, respectively. Half the amounts of plasmid and jetPRIME were used in each round for co-overexpression experiments. At 18 hours after transfection, EGF (Peprotech) was added to a final concentration of 100 ng/ml. At 20 minutes before a given EGF stimulation time point, 5-iodo-deoxycytidine (IdU) was added to the medium at the final concentration of 10 µM. At 2 minutes before a given EGF stimulation

time point, medium was replaced by 1X TrypLE to induce cell detachment. At the time point, paraformaldehyde (PFA, from Electron Microscopy Sciences) was added to the cell suspension to a final percentage of 1.6%, and cells were incubated at room temperature for 10 minutes. If EGF stimulation was not necessary in the experiment, cells were directly harvested and crosslinked with PFA.

***A375 cell transfection and vemurafenib treatment***

A375 cells were seeded at the density of 0.15 million per well in 6-well plates. At 24 hours after seeding, transfection was performed using 2 µg plasmid and 4 µl of X-treme GENE HP reagent (06 366 236 001, Roche) per well with the standard protocol provided by the manufacturer. Pre-diluted vemurafenib (10 mM/mL in DMSO, Selleckchem) was added to the cells 18 hours after transfection, at the final concentration of 1 µM. The same volume of DMSO was added to the control samples. After 2 days, cells were labeled with IdU during 20-minute incubation and subsequently harvested by 5-minute TrypLE digestion and 10-minute PFA crosslinking as described above.

***Methanol permeabilization***

Crosslinked cells were washed twice with cell staining media (CSM, PBS with 0.5% BSA). After centrifugation, ice-cold methanol was used to resuspend the cells, followed by 10-minute permeabilization on ice or for long-term storage at -80 °C.

***Antibody conjugation***

The MaxPAR antibody conjugation kit (Fluidigm) was used to generate isotope-labeled antibodies using the manufacturer's standard protocol. After conjugation, the antibody yield was determined based on absorbance of 280 nm. Candor PBS Antibody Stabilization solution (Candor Bioscience GmbH) was used to dilute antibodies for long-term storage at 4 °C.

***Barcoding and staining protocol***

Formalin-crosslinked and methanol-permeabilized cells were washed three times with CSM and once with PBS. Cells were incubated in PBS containing barcoding reagents of $^{89}$Y (100 nM), $^{103}$Rh (2 µM), $^{105}$Pd (100 nM), $^{106}$Pd (100 nM), $^{108}$Pd (100 nM), $^{110}$Pd (100 nM), $^{113}$In (200 nM), $^{115}$In (100 nM), and $^{209}$Bi (20 nM) for 30 minutes at room temperature and then washed three times with CSM. Barcoded cells were then pooled and stained with the metal-conjugated

antibody mix (Table S1) at room temperature for 1 hour. The antibody mix was removed by washing cells three times with CSM and once with PBS. For DNA staining, iridium-containing intercalator (Fluidigm) diluted in PBS with 1.6% PFA was incubated with the cells at 4 °C overnight. On the day of the measurement, the intercalator solution was removed, and cells were washed with CSM, PBS, and $ddH_2O$. After the last washing step, cells were resuspended in $ddH_2O$ and filtered through a 70-µm strainer.

***Mass cytometry analysis***

EQ™ Four Element Calibration Beads (Fluidigm) were added to cell suspensions in a 1:10 ratio (v/v). Samples were analyzed on a CyTOF2 (Fluidigm). The manufacturer's standard operation procedures were used for acquisition at a cell rate of ~500 cells per second. After the acquisition, all FCS files from the same barcoded sample were concatenated (Bodenmiller et al., 2012). Data were then normalized, and bead events were removed (Finck et al., 2013) before doublet removal and de-barcoding of cells into their corresponding wells using a doublet-filtering scheme and single-cell deconvolution algorithm (Zunder et al., 2015). Subsequently, data was processed using Cytobank (http://www.cytobank.org). Additional gating on the DNA channels ($^{191}$Ir and $^{193}$Ir) and $^{139}$La/$^{141}$Pr was used to remove remained doublets, debris and contaminating particulate.

***Data preprocessing and BP-$R^2$ analysis***

*Bi-axis scatter plots*

Bi-axis scatter plots were generated in Cytobank (http://www.cytobank.org/).

*Data preprocessing*

Raw data was transformed using the inverse hyperbolic sine transform with a cofactor of 5:

$$data = arsinh(data_{raw}/5)$$

Except where use of raw data values is specifically noted, all visualizations and analyses were performed using transformed data.

*Data binning*

For data binning, the range between the lower and upper 2.5% of observations was divided into ten equal bins $bin_1$,...,$bin_{10}$. The observations in the lower and upper 2.5% were assigned to the lowest and highest bins, respectively. In order to be able to compare expression levels between samples within a time course experiment, all observations of the time course were pooled to determine the binning.

*BP-R2*

BP-$R^2$ analysis is described in Lun et al., 2017 (https://github.com/BodenmillerGroup/Adnet). In brief, the median of a measured marker ($\tilde{y}_i$) was calculated for each bin $i$. Additionally, the overall mean of the medians of all the 10 bins ($\mu_{\tilde{y}}$) was calculated (bins with less than 25 cells were discarded). Then, for each bin, we computed the sum of squared deviations from the bin medians and the sum of squared deviations from the overall mean of medians. These values were summed over all bins and the BP-$R^2$ was defined as one minus the ratio between them:

$$R^2_{BP} = 1 - \frac{\sum_{i=1}^{nbins} \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \tilde{y}_i)^2}{\sum_{i=1}^{nbins} \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \mu_{\tilde{y}})^2}$$

*Threshold determination*

Following the method described in Lun et al., 2017, we chose the maximum BP-$R^2$ among all the 108 control samples (FLAG-GFP overexpression and untransfected cells) as a cutoff. Relationships that had a BP-$R^2$ higher than this threshold were considered as sufficiently strong to be of interest.

*Signed-BP-R2*

The relationship strengths calculated as BP-$R^2$ were mostly positive, with a few exceptions of negative BP-$R^2$ values mostly from the cell cycle marker IdU, due to bimodality. These rare and weak negative BP-$R^2$ values were considered as negligible and were therefore assigned to 0. This allowed the integration of signaling relationship directionalities, determined by Spearman

correlation of bin medians ($\rho_{bin}$), with the relationship strengths ($R^2_{BP}$). The signed-BP-$R^2$ score ($R^2_{signed-BP}$) was calculated as:

$$R^2_{Signed-BP} = \begin{cases} R^2_{BP}, & \rho_{bin} \geq 0 \\ -R^2_{BP}, & \rho_{bin} < 0 \end{cases}$$

***Hierarchical clustering***

Hierarchical clustering was performed for kinases and phosphatases on their abundance-dependent signaling relationships, as signed-BP-$R^2$, to all measured phosphorylation sites with and without 10-minute EGF stimulation. Ward's method and Euclidean distances (Ward, 1963) were used for the clustering, and the hierarchical tree was cut at the height of 5 to obtain 10 clusters of kinases and phosphatases as shown in Figure S1C.

***t-SNE analysis***

t-SNE analysis was performed with the Package Rtsne in R.

***Functional enrichment and association analysis using STRING database***

The functional enrichment and interaction enrichment analyses were performed using the STRING database v10.5 (Szklarczyk et al., 2017). All the kinases and phosphates tested were mapped to STRING protein name-space establishing the background protein set for the further analysis. The functional enrichment p-values were corrected using Benjamini and Hochberg method (Benjamini and Hochberg, 1995) (the detailed description of the statistical methods can be found in Franceschini et al., 2012). To test whether the functional signal within the clusters arises exclusively from a homology between the proteins, the homologous proteins were grouped together into one node, and, therefore, the proteins that exhibited high or medium homology did not contribute independently to the enrichment functional term count. In order to form the grouped representation of the STRING network the single-linkage clustering method was applied to the homology relationships between the proteins in which neighbors were defined as having a self-normalized bit score (BLAST bit score of alignment between the two proteins divided by the bit score of self-alignment of shorter of the two proteins) equal to or higher than 0.2. For each functional term the grouped node contributed to the enrichment count when one or more of the proteins forming the group were annotated with the term in question. This process was applied to both the clusters and the background separately to ensure that for

the groups in which proteins were shared between the cluster and the background the functional term was counted in both sets.

***Shortest signed directed path analysis using OmniPath***

The pathway analysis was performed for all signaling relationships between overexpressed POIs and measured phosphorylation sites using OmniPath (http://omnipathdb.org/), a collection of literature curated human signaling pathways integrated from 25 databases (pathway databases: TRIP, SPIKE, SignaLink3, Guide2Pharma, CA1, ARN, NRF2ome, Macrophage, DeathDomain, PDZBase, Signor; interaction databases: BioGRID, CancerCellMap, MPPI, DIP, InnateDB, MatrixDB; PTM databases: PhosphoSite, DEPOD, LMPID, phosphoELM, ELM, DOMINO, dbPTM, HPRD-phos) (Türei et al., 2016). The shortest path was determined based on based on Breadth-First Search methods, computed through a Python module called pyPath (Türei et al., 2016).

***Selection of strong signaling dynamic influencing POIs***

For each pair of signaling relationships between an overexpressed POI and a measured phosphorylation site, the delta BP-$R^2$ score was calculated as the signed-BP-$R^2$ value with 10 minute EGF stimulation minus the signed-BP-$R^2$ value in unstimulated cells. We selected the top 10 POIs with a positive difference in signed-BP-$R^2$, the top 10 POIs with negative difference in signed-BP-$R^2$, the top 20 POIs with the number of signaling relationships in the 99 percentile of the difference in signed-BP-$R^2$, and the 10 central signaling dynamic regulators in the MAPK/ERK and AKT pathways known from the literature (Steelman and Chappell, 2011). Some POIs were in more than one set, so this resulted in 39 kinases and phosphatases.

***Signaling dynamics analysis***

The signaling dynamics analysis was adapted from our previous methods (Lun et al., 2017) that characterize two signaling aspects, the protein abundance-dependent signaling amplitudes and peak times. For amplitude analysis, the fold change of median phosphorylation abundance for each bin in EGF-stimulated samples over the corresponding bin of the unstimulated sample (EGF 0 min) was calculated using the raw count. The amplitude for each bin was identified as the maximal fold change over all time points. Amplitude ratios between the second highest and the second lowest bin amplitudes were computed for all samples, and the highest amplitude ratio in all FLAG-GFP overexpression and untransfected controls was used to determine the

cutoff for robust and strong abundance-dependent changes. To study protein abundance-dependent signaling peak times, the time point after EGF addition with maximal amplitude for each bin was defined as peak time. We calculated the maximal median variation over all bins for each individual sample and selected pairs of relationships with bin variations (at least in one time point) higher than all those in control samples. These were defined as peak time shifts.

## Figures

Figure 1

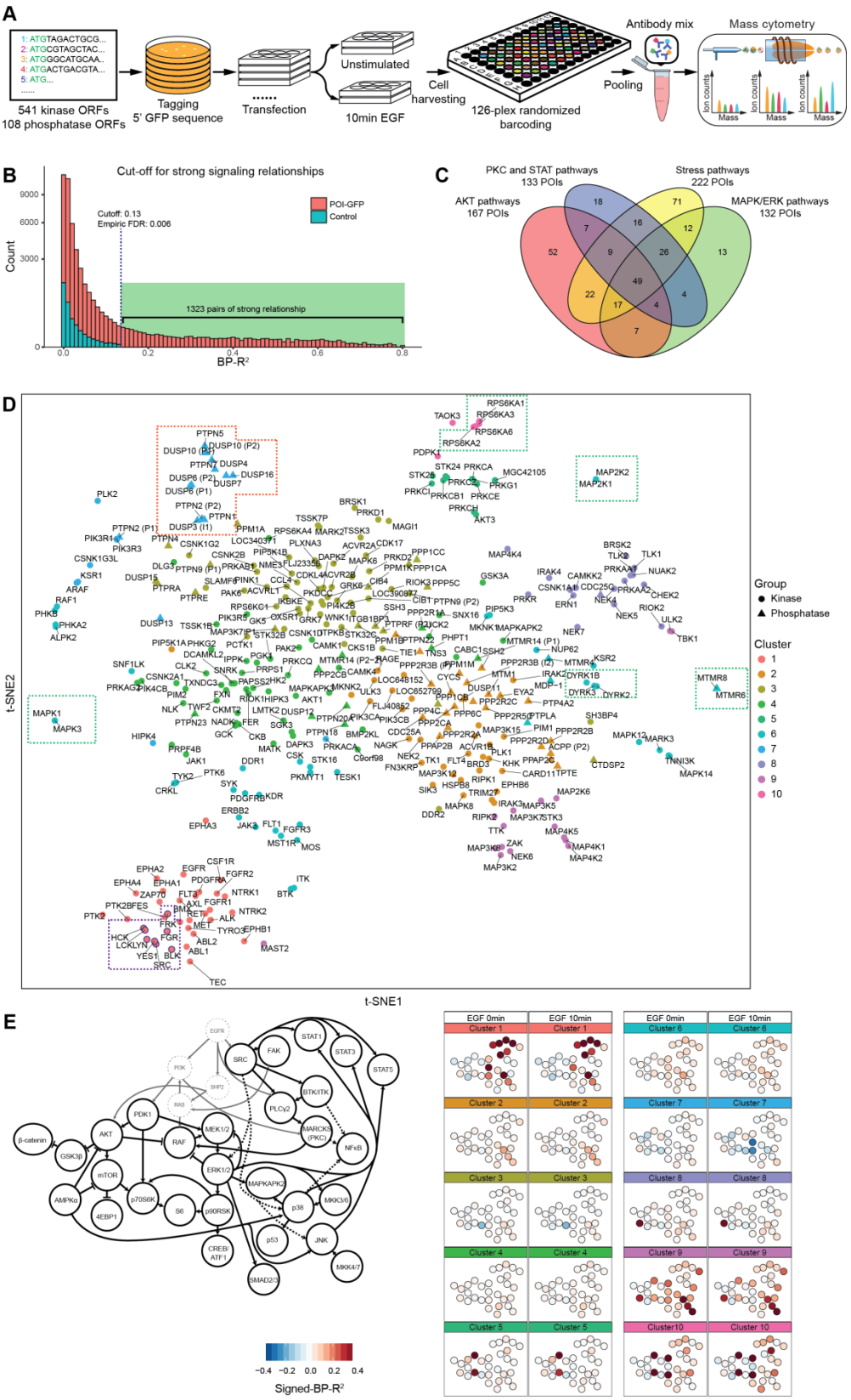


Figure 2

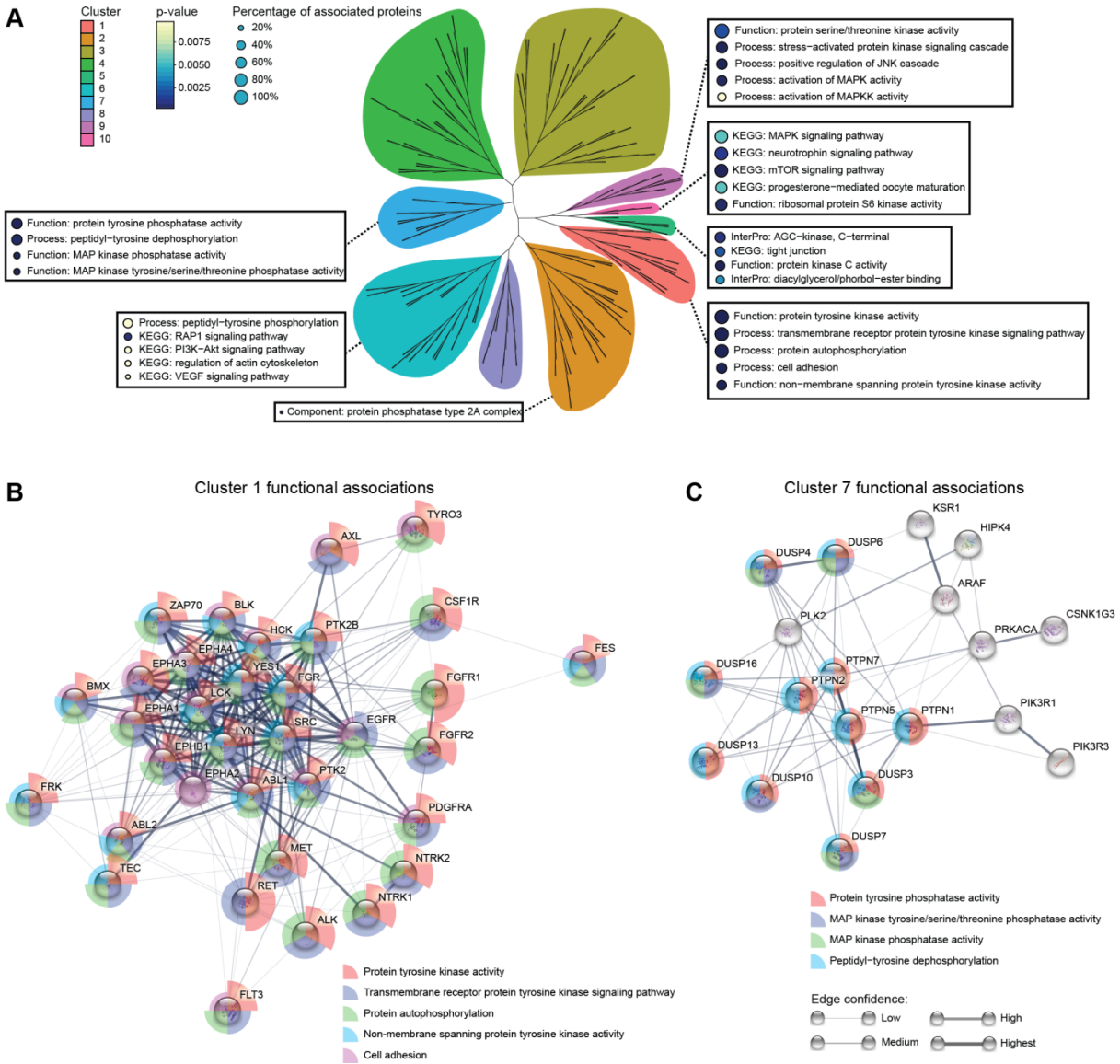

A
Cluster
1
2
3
4
5
6
7
8
9
10
p-value
0.0075
0.0050
0.0025
Percentage of associated proteins
20%
40%
60%
80%
100%
Function: protein serine/threonine kinase activity
Process: stress−activated protein kinase signaling cascade
Process: positive regulation of JNK cascade
Process: activation of MAPK activity
Process: activation of MAPKK activity
KEGG: MAPK signaling pathway
KEGG: neurotrophin signaling pathway
KEGG: mTOR signaling pathway
KEGG: progesterone−mediated oocyte maturation
Function: ribosomal protein S6 kinase activity
InterPro: AGC−kinase, C−terminal
KEGG: tight junction
Function: protein kinase C activity
InterPro: diacylglycerol/phorbol−ester binding
Function: protein tyrosine kinase activity
Process: transmembrane receptor protein tyrosine kinase signaling pathway
Process: protein autophosphorylation
Process: cell adhesion
Function: non−membrane spanning protein tyrosine kinase activity
Function: protein tyrosine phosphatase activity
Process: peptidyl−tyrosine dephosphorylation
Function: MAP kinase phosphatase activity
Function: MAP kinase tyrosine/serine/threonine phosphatase activity
Process: peptidyl−tyrosine phosphorylation
KEGG: RAP1 signaling pathway
KEGG: PI3K−Akt signaling pathway
KEGG: regulation of actin cytoskeleton
KEGG: VEGF signaling pathway
Component: protein phosphatase type 2A complex
B
Cluster 1 functional associations
TYRO3
AXL
ZAP70
BLK
CSF1R
HCK
PTK2B
FES
EPHA4
EPHA3
YES1
FGR
FGFR1
BMX
LCK
EPHA1
SRC
EGFR
LYN
FGFR2
EPHB1
EPHA2
ABL1
PTK2
FRK
PDGFRA
ABL2
MET
NTRK2
TEC
RET
NTRK1
ALK
FLT3
Protein tyrosine kinase activity
Transmembrane receptor protein tyrosine kinase signaling pathway
Protein autophosphorylation
Non−membrane spanning protein tyrosine kinase activity
Cell adhesion
C
Cluster 7 functional associations
KSR1
HIPK4
DUSP4
DUSP6
ARAF
PLK2
CSNK1G3
PRKACA
DUSP16
PTPN7
PTPN2
PTPN5
PTPN1
PIK3R1
DUSP13
PIK3R3
DUSP10
DUSP3
DUSP7
Protein tyrosine phosphatase activity
MAP kinase tyrosine/serine/threonine phosphatase activity
MAP kinase phosphatase activity
Peptidyl−tyrosine dephosphorylation
Edge confidence:
Low
High
Medium
Highest

Figure 3

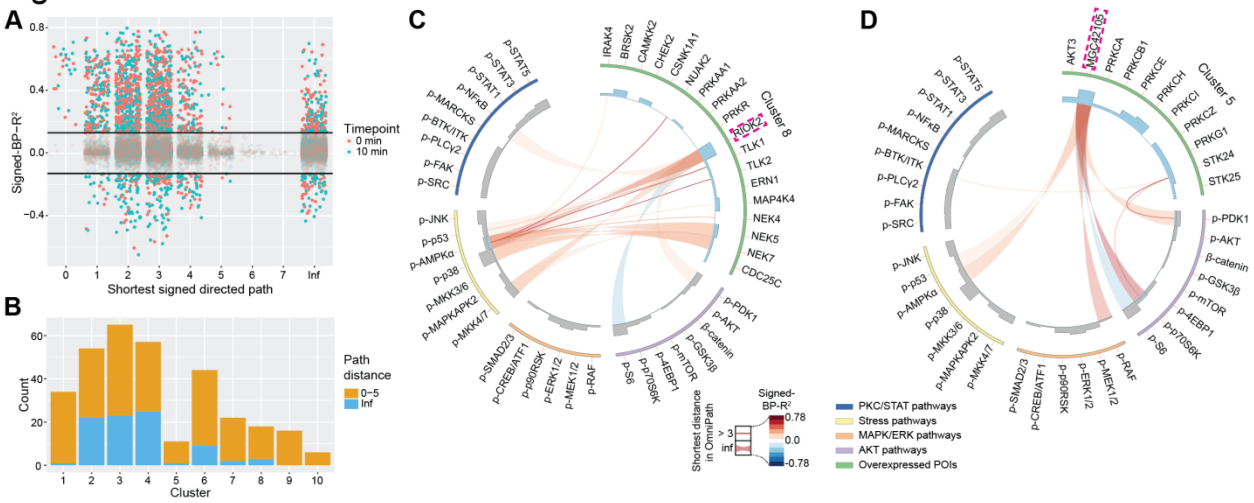

A
Signed-BP−R²
Shortest signed directed path
Timepoint
0 min
10 min
B
Count
Cluster
Path distance
0–5
Inf
C
Cluster 8
IRAK4
BRSK2
CAMKK2
CHEK2
CSNK1A1
NUAK2
PRKAA1
PRKAA2
PRKR
NLRC2
TLK1
TLK2
ERN1
MAP4K4
NEK4
NEK5
NEK7
CDC25C
p-PDK1
p-AKT
p-catenin
p-GSK3β
p-mTOR
p-4EBP1
p-p70S6K
p-S6
p-RAF
p-MEK1/2
p-ERK1/2
p-p90RSK
p-CREB/ATF1
p-SMAD2/3
p-MKK4/7
p-MAPKAPK2
p-MKK3/6
p-p38
p-AMPKα
p-p53
p-JNK
p-SRC
p-FAK
p-PLCγ2
p-BTK/ITK
p-MARCKS
p-NFκB
p-STAT1
p-STAT3
p-STAT5
Shortest distance in OmniPath
> 3
inf
Signed-BP-R²
0.78
0.0
-0.78
D
Cluster 5
AKT3
MGST2
PRKCA
PRKCB1
PRKCE
PRKCH
PRKCI
PRKCZ
PRKG1
STK24
STK25
PKC/STAT pathways
Stress pathways
MAPK/ERK pathways
AKT pathways
Overexpressed POIs

Figure 4

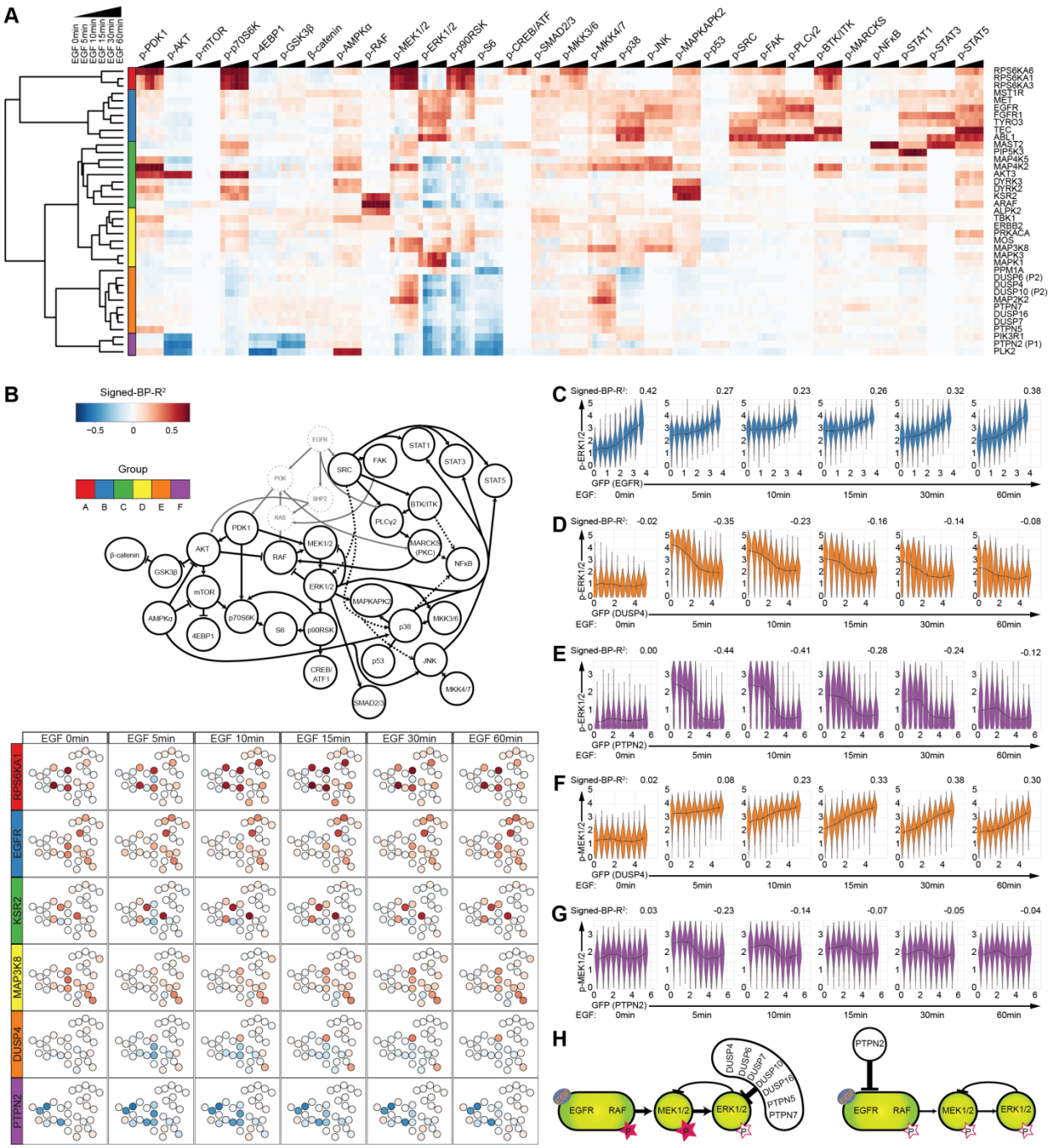
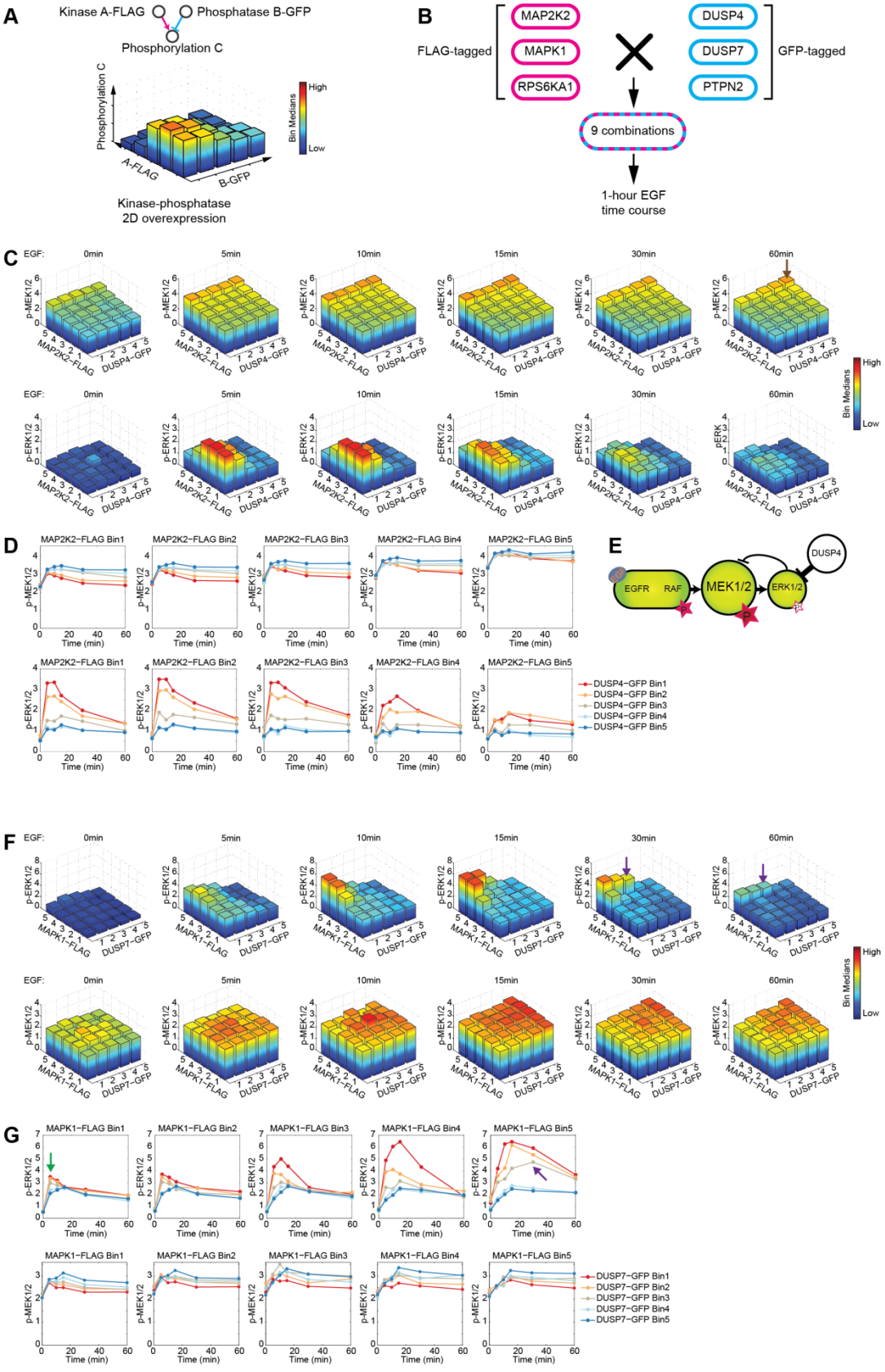
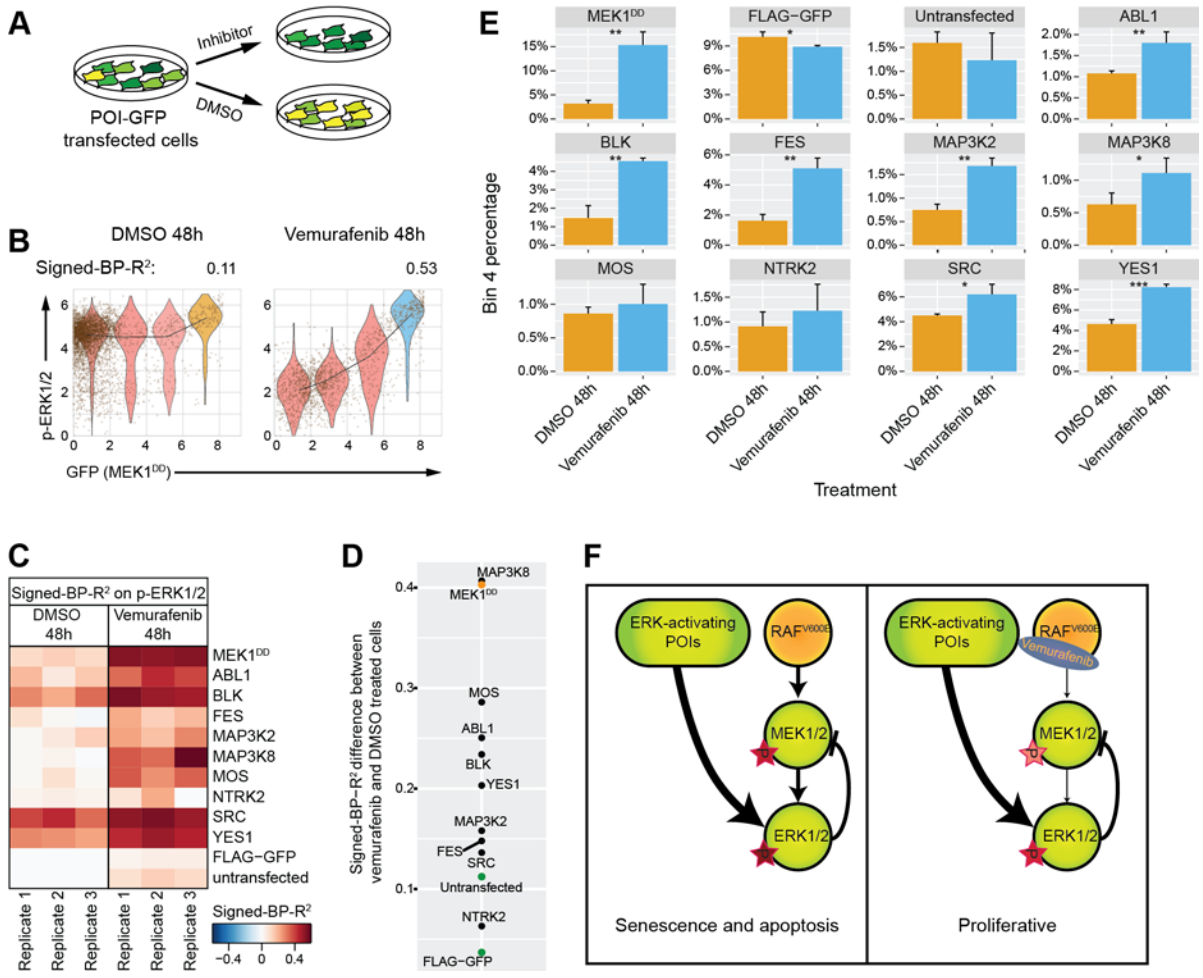

A
B
C
D
E
F
G
H
Signed-BP-R²
Group
EGF 0min
EGF 5min
EGF 10min
EGF 15min
EGF 30min
EGF 60min
p-PDK1
p-AKT
p-mTOR
p-p70S6K
p-4EBP1
p-GSK3β
β-catenin
p-AMPKα
p-RAF
p-MEK1/2
p-ERK1/2
p-p90RSK
p-S6
p-CREB/ATF
p-SMAD2/3
p-MKK3/6
p-MKK4/7
p-p38
p-JNK
p-MAPKAPK2
p-p53
p-SRC
p-FAK
p-PLCγ2
p-BTK/ITK
p-MARCKS
p-NFκB
p-STAT1
p-STAT3
p-STAT5
RPS6KA6
RPS6KA1
RPS6KA3
MST1R
MET
EGFR
FGFR1
TYRO3
TEC
ABL1
MAST2
PIP5K3
MAP4K5
MAP4K2
AKT3
DYRK3
DYRK2
KSR2
ARAF
ALPK2
TBK1
ERBB2
PRKACA
MOS
MAP3K8
MAPK3
MAPK1
PPM1A
DUSP6 (P2)
DUSP4
DUSP10 (P2)
MAP2K2
PTPN7
DUSP16
DUSP7
PTPN5
PIK3R1
PTPN2 (P1)
PLK2
EGFR
PI3K
SHP2
RAS
SRC
FAK
STAT1
STAT3
STAT5
BTK/ITK
PLCγ2
MARCKS (PKC)
NFκB
PDK1
AKT
β-catenin
GSK3β
RAF
MEK1/2
mTOR
ERK1/2
AMPKα
4EBP1
p70S6K
S6
p90RSK
MAPKAPK2
p38
MKK3/6
p53
JNK
CREB/ ATF1
MKK4/7
SMAD2/3
RPS6KA1
EGFR
KSR2
MAP3K8
DUSP4
PTPN2
Signed-BP-R²: 0.42 0.27 0.23 0.26 0.32 0.38
p-ERK1/2
GFP (EGFR)
EGF: 0min 5min 10min 15min 30min 60min
Signed-BP-R²: -0.02 -0.35 -0.23 -0.16 -0.14 -0.08
GFP (DUSP4)
Signed-BP-R²: 0.00 -0.44 -0.41 -0.28 -0.24 -0.12
GFP (PTPN2)
Signed-BP-R²: 0.02 0.08 0.23 0.33 0.38 0.30
p-MEK1/2
Signed-BP-R²: 0.03 -0.23 -0.14 -0.07 -0.05 -0.04
EGFR
RAF
MEK1/2
ERK1/2
DUSP4
DUSP6
DUSP7
DUSP13
DUSP16
PTPN5
PTPN7
PTPN2

Figure 5


A
Kinase A-FLAG
Phosphatase B-GFP
Phosphorylation C
Phosphorylation C
A-FLAG
B-GFP
Bin Medians
High
Low
Kinase-phosphatase
2D overexpression
B
FLAG-tagged
MAP2K2
MAPK1
RPS6KA1
DUSP4
DUSP7
PTPN2
GFP-tagged
9 combinations
1-hour EGF
time course
C
EGF:
0min
5min
10min
15min
30min
60min
p-MEK1/2
p-ERK1/2
pERK
MAP2K2-FLAG
DUSP4-GFP
Bin Medians
High
Low
D
MAP2K2-FLAG Bin1
MAP2K2-FLAG Bin2
MAP2K2-FLAG Bin3
MAP2K2-FLAG Bin4
MAP2K2-FLAG Bin5
p-MEK1/2
p-ERK1/2
Time (min)
DUSP4-GFP Bin1
DUSP4-GFP Bin2
DUSP4-GFP Bin3
DUSP4-GFP Bin4
DUSP4-GFP Bin5
E
EGFR
RAF
MEK1/2
ERK1/2
DUSP4
P
F
EGF:
0min
5min
10min
15min
30min
60min
p-ERK1/2
p-MEK1/2
MAPK1-FLAG
DUSP7-GFP
Bin Medians
High
Low
G
MAPK1-FLAG Bin1
MAPK1-FLAG Bin2
MAPK1-FLAG Bin3
MAPK1-FLAG Bin4
MAPK1-FLAG Bin5
p-ERK1/2
p-MEK1/2
Time (min)
DUSP7-GFP Bin1
DUSP7-GFP Bin2
DUSP7-GFP Bin3
DUSP7-GFP Bin4
DUSP7-GFP Bin5

# Figure 6


A
Inhibitor
DMSO
POI-GFP
transfected cells
B
DMSO 48h
Vemurafenib 48h
Signed-BP-R²:
0.11
0.53
p-ERK1/2
GFP (MEK1DD)
E
MEK1DD
FLAG−GFP
Untransfected
ABL1
BLK
FES
MAP3K2
MAP3K8
MOS
NTRK2
SRC
YES1
Bin 4 percentage
DMSO 48h
Vemurafenib 48h
Treatment
C
Signed-BP-R² on p-ERK1/2
DMSO 48h
Vemurafenib 48h
MEK1DD
ABL1
BLK
FES
MAP3K2
MAP3K8
MOS
NTRK2
SRC
YES1
FLAG−GFP
untransfected
Replicate 1
Replicate 2
Replicate 3
Signed-BP-R²
−0.4
0
0.4
D
Signed-BP-R² difference between vemurafenib and DMSO treated cells
MAP3K8
MEK1DD
MOS
ABL1
BLK
YES1
MAP3K2
FES
SRC
Untransfected
NTRK2
FLAG−GFP
F
ERK-activating POIs
RAFV600E
MEK1/2
ERK1/2
Vemurafenib
Senescence and apoptosis
Proliferative

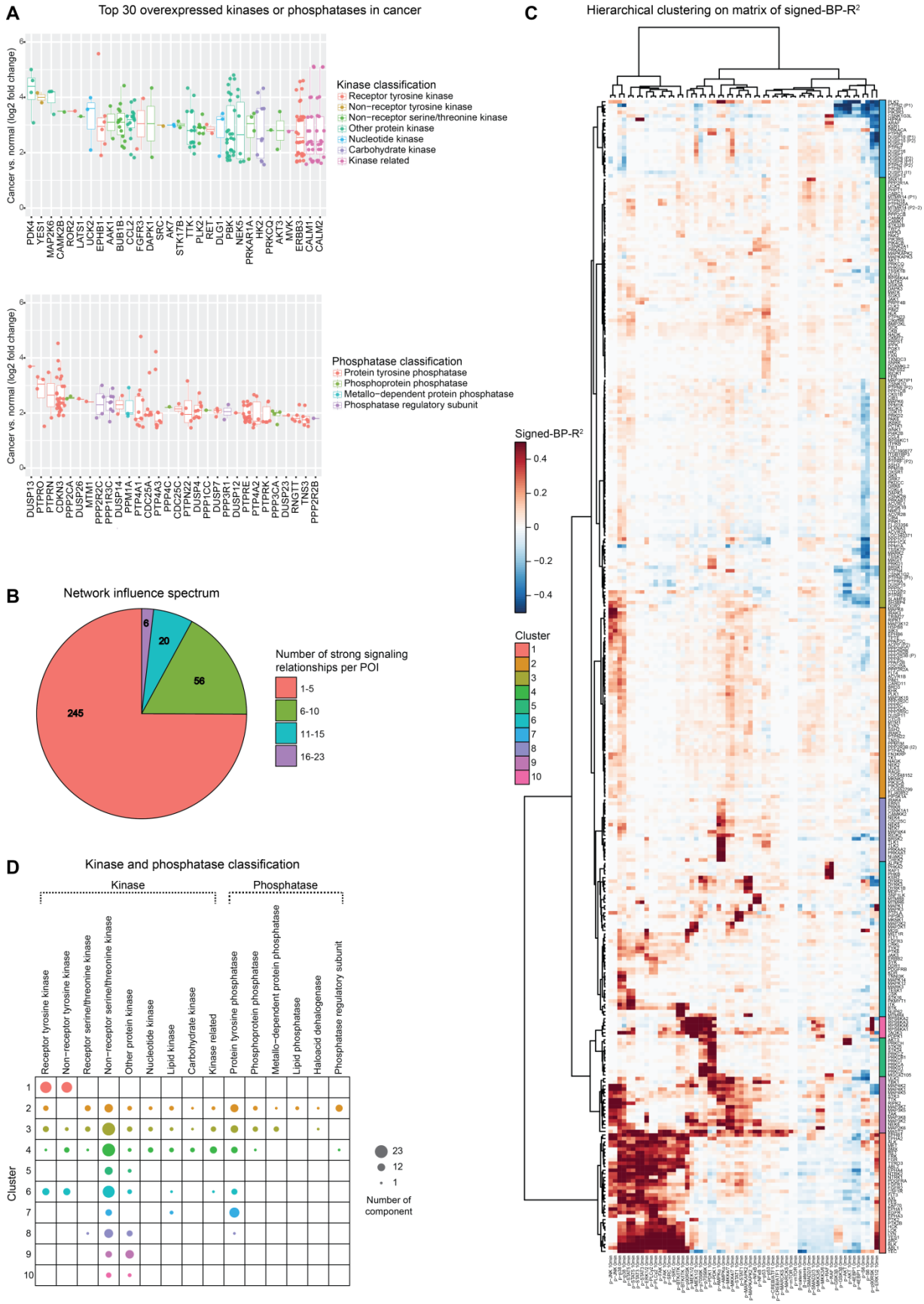

A
Top 30 overexpressed kinases or phosphatases in cancer
Cancer vs. normal (log2 fold change)
Kinase classification
Receptor tyrosine kinase
Non−receptor tyrosine kinase
Non−receptor serine/threonine kinase
Other protein kinase
Nucleotide kinase
Carbohydrate kinase
Kinase related
PDK4
YES1
MAP2K6
CAMK2B
PRKCB
LATS1
UCK2
EPHB1
AAK1
BUB1B
CCL2
FGFR3
DAPK1
SRC
AK7
STK17B
TTK
PLK2
RET
DLG1
PBK
NEK6
PRKAR1A
HK2
PRKCQ
AKT3
MVK
ERBB3
CALM1
CALM2
Phosphatase classification
Protein tyrosine phosphatase
Phosphoprotein phosphatase
Metallo−dependent protein phosphatase
Phosphatase regulatory subunit
DUSP13
PTPRO
PTPRN
CDKN3
PPP2CA
DUSP26
MTM1
PPP2R2C
PPP1R3C
DUSP14
PPM1A
PTP4A1
CDC25A
PTP4A3
PPP4C
CDC25C
PTPN22
DUSP4
PPP1CC
DUSP7
PPP3R1
DUSP12
PTPRE
PTP4A2
PTPRK
PPP3CA
DUSP23
RNGTT
TNS3
PPP2R2B
B
Network influence spectrum
Number of strong signaling relationships per POI
1-5
6-10
11-15
16-23
245
56
20
6
C
Hierarchical clustering on matrix of signed-BP-R²
Signed-BP-R²
Cluster
D
Kinase and phosphatase classification
Kinase
Phosphatase
Receptor tyrosine kinase
Non−receptor tyrosine kinase
Receptor serine/threonine kinase
Non−receptor serine/threonine kinase
Other protein kinase
Nucleotide kinase
Lipid kinase
Carbohydrate kinase
Kinase related
Protein tyrosine phosphatase
Phosphoprotein phosphatase
Metallo−dependent protein phosphatase
Lipid phosphatase
Haloacid dehalogenase
Phosphatase regulatory subunit
Cluster
Number of component

**Figure S1. Kinase and phosphatase overexpression impacts cell signaling. A**, Top 30 overexpressed kinases or phosphatases in cancers or cancer cell lines compared to normal cells, using data from the Expression Atlas (Petryszak et al., 2016). **B**, Pie chart shows the number of POIs that have strong (BP-$R^2$ > 0.13) overexpression-induced signaling relationships to the measured phosphorylation sites. Three-fourths of the POIs modulated one to five phosphorylation sites. **C**, Heat map shows the signed-BP-$R^2$ of all 327 EGFR network-influencing POIs (i.e., POIs with at least one strong overexpression-induced signaling relationship, BP-$R^2$ > 0.13) to every phosphorylation site with or without 10-minute EGF stimulation. Hierarchical clustering identified 10 groups of kinases or phosphatases with the similar network-influencing properties. **D**, Corresponding components between identified hierarchical clusters (the same cluster color code as in panel (**C**) and kinase/phosphatase classifications based on the sequence of protein catalytic domains.

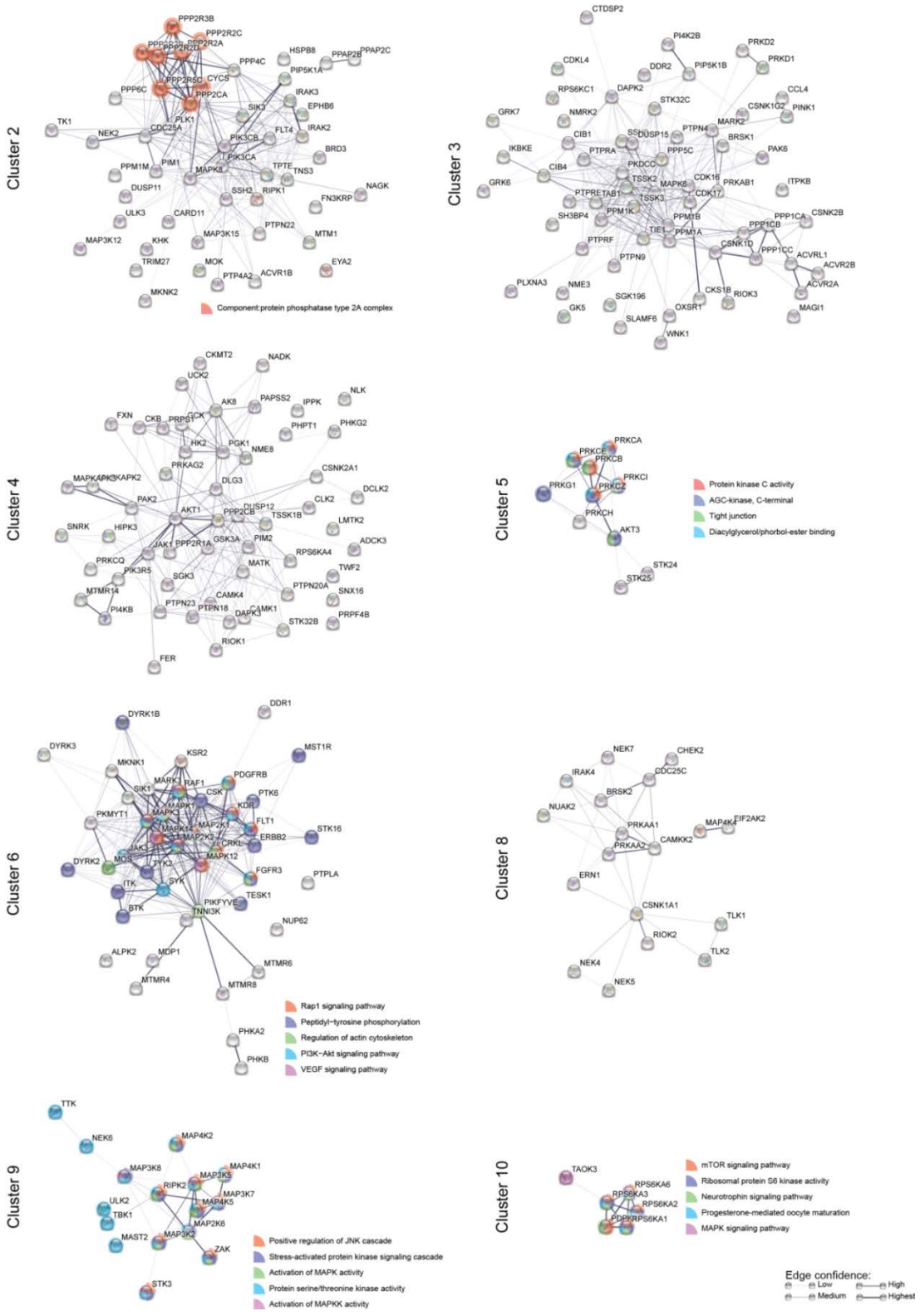



**Figure S2. Functional association analysis.** Functional association analysis was performed for clusters 2, 3, 4, 5, 6, 8, 9, and 10. Edges with confidence above 0.2 are shown in the network. Functional enrichments are color-coded and labeled on the nodes.

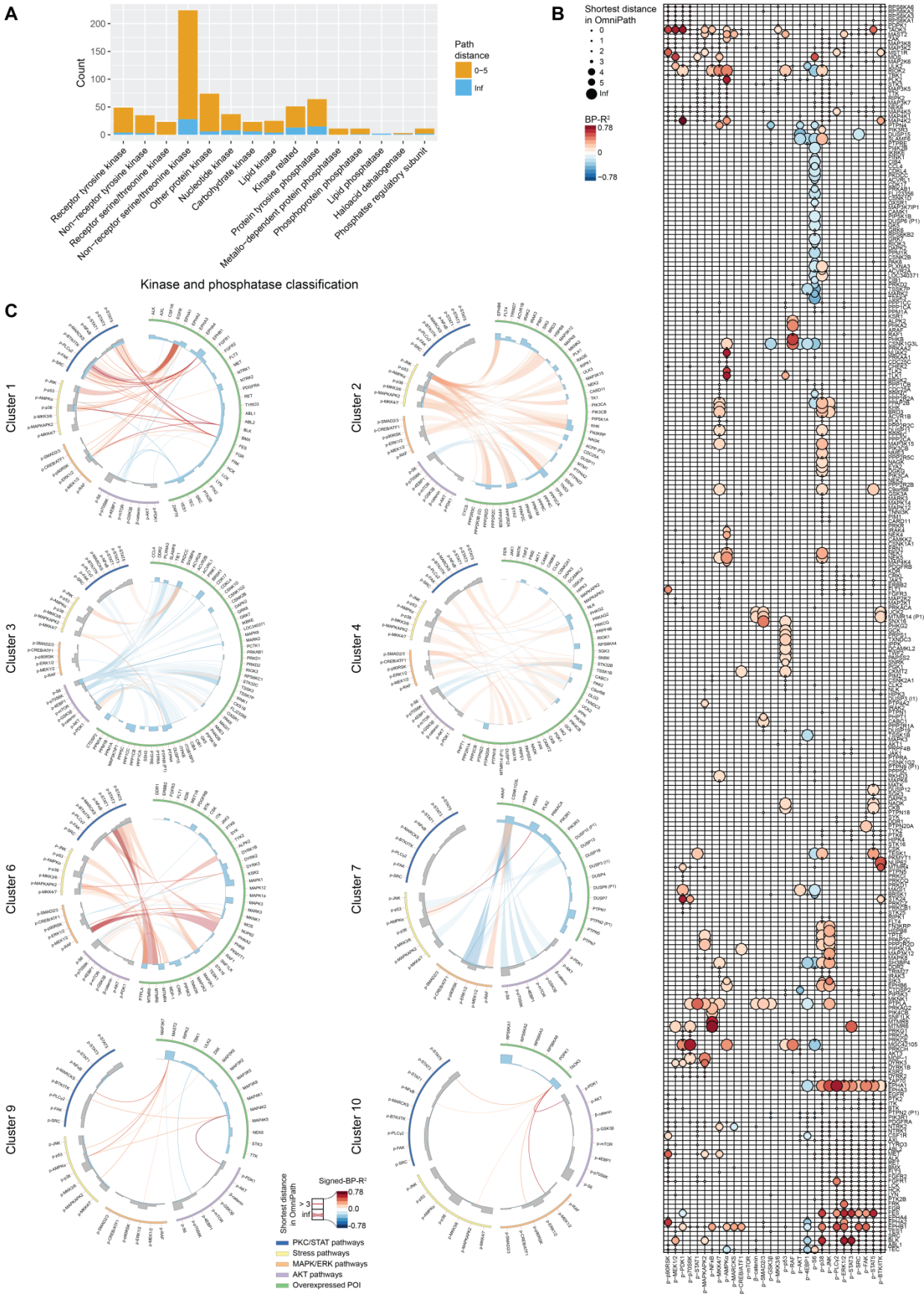

A
Count
200
150
100
50
0
Path distance
0–5
Inf
Receptor tyrosine kinase
Non−receptor tyrosine kinase
Receptor serine/threonine kinase
Non−receptor serine/threonine kinase
Other protein kinase
Nucleotide kinase
Carbohydrate kinase
Lipid kinase
Kinase related
Protein tyrosine phosphatase
Metallo−dependent protein phosphatase
Phosphoprotein phosphatase
Lipid phosphatase
Haloacid dehalogenase
Phosphatse regulatory subunit
Kinase and phosphatase classification
B
Shortest distance in OmniPath
0
1
2
3
4
5
Inf
BP-R²
0.78
0
−0.78
C
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 6
Cluster 7
Cluster 9
Cluster 10
Signed-BP-R²
0.78
0.0
-0.78
> 3
inf
Shortest distance in OmniPath
PKC/STAT pathways
Stress pathways
MAPK/ERK pathways
AKT pathways
Overexpressed POI

**Figure S3. Novel signaling relationships detected by comparison with the OmniPath database. A**, Counts of strong (BP-$R^2$ > 0.13) abundance-dependent signaling relationships with shortest signed directed path length of 0-5 or with infinite path length for each sequenced-based kinase and phosphatase classification. **B**, Abundance-dependent signaling relationships are indicated as circles with circle size indicating the shortest signed directed path length and color the signed-BP-$R^2$ score. **C**, Shortest signed directed path length for each identified signaling relationships shown in Circos plots for the clusters of 1, 2, 3, 4, 6, 7, 9, and 10.

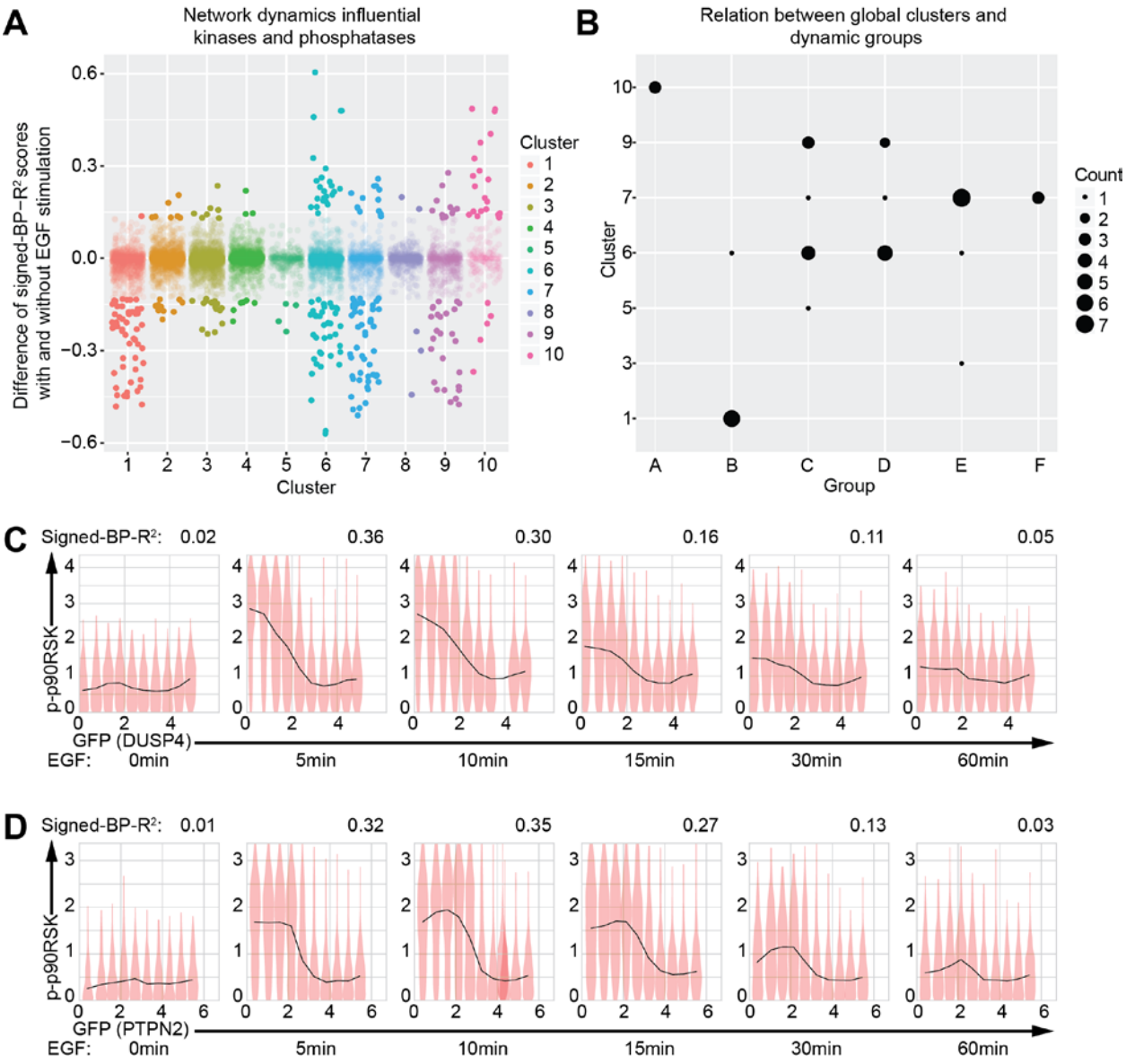



**Figure S4. Analysis of kinase and phosphatase overexpression-dependent signaling dynamics in EGF stimulation time course. A**, Variances of signed-BP-$R^2$ between cells with or without 10-minute EGF treatment are plotted for all identified clusters in the global analysis. **B**, Correspondence analysis shows to which 10 clusters as derived from the kinome and phosphatome analysis, the six dynamic signaling groups belong to. **C-D**, Violin plots show cell distribution in each of the 10 bins based on GFP-tagged POI expression level for (**C**) DUSP4-GFP to p-p90RSK, and (**D**) PTPN2-GFP to p-p90RSK over the 1-hour EGF stimulation time course. Medians of all 10 bins are connected to indicate the shape of signaling relationships (black lines) with the relationship strength quantified by signed-BP-$R^2$, as shown on top of each individual plot.

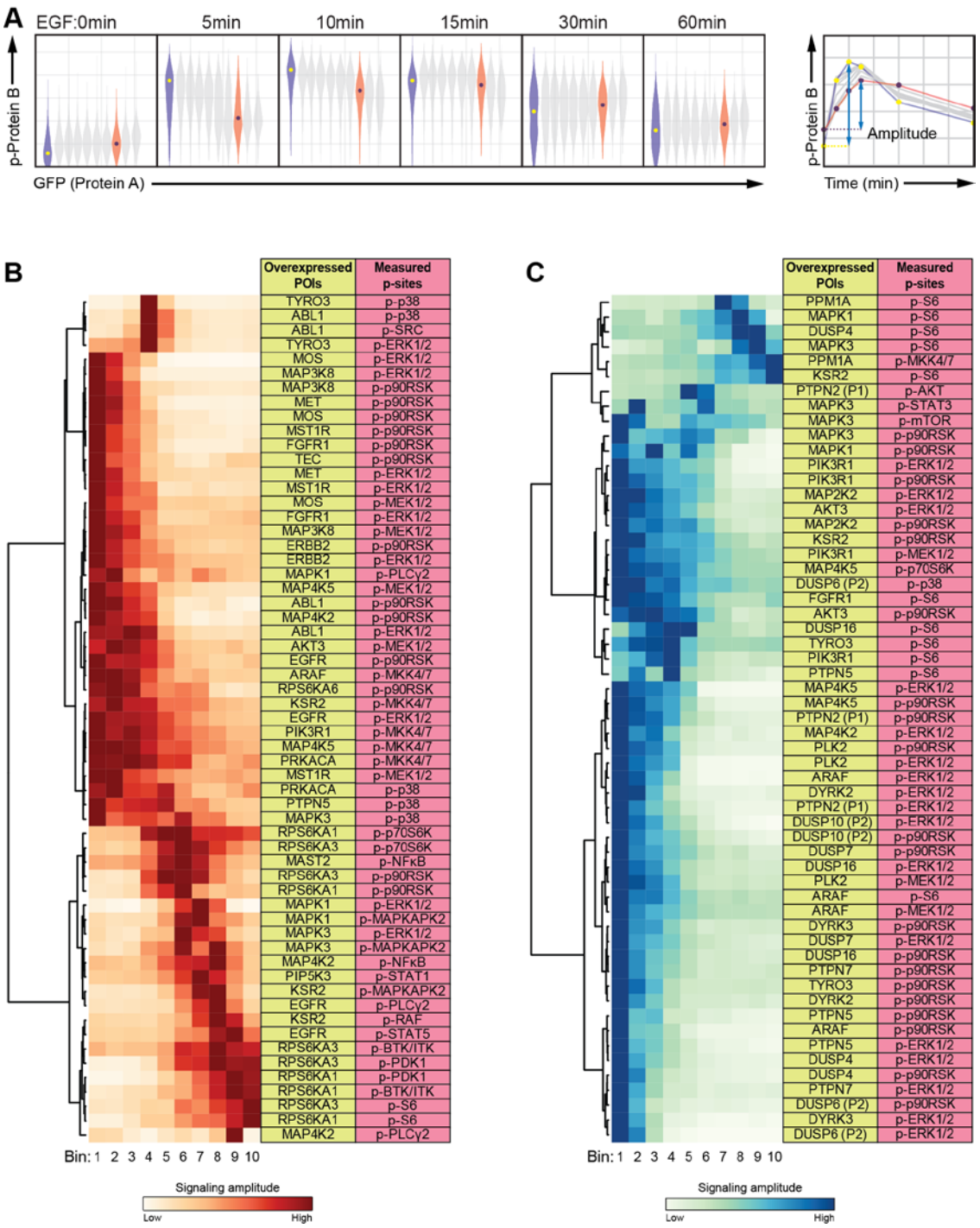



**Figure S5. Protein abundance-modulated signaling amplitudes. A**, Schematic plots of amplitude analysis. The abundance levels of the overexpressed GFP-POI were split into ten bins (left). Median phosphorylation in each bin over the 1-hour EGF stimulation time course are plotted to visualize abundance dependency of signaling amplitudes (right). **B-C**, Heat maps show pairs of signaling relationships with protein abundance-influenced signaling amplitudes with (**B**) positive relationships and (**C**) negative relationships determined by the summed Spearman correlation over all time points.

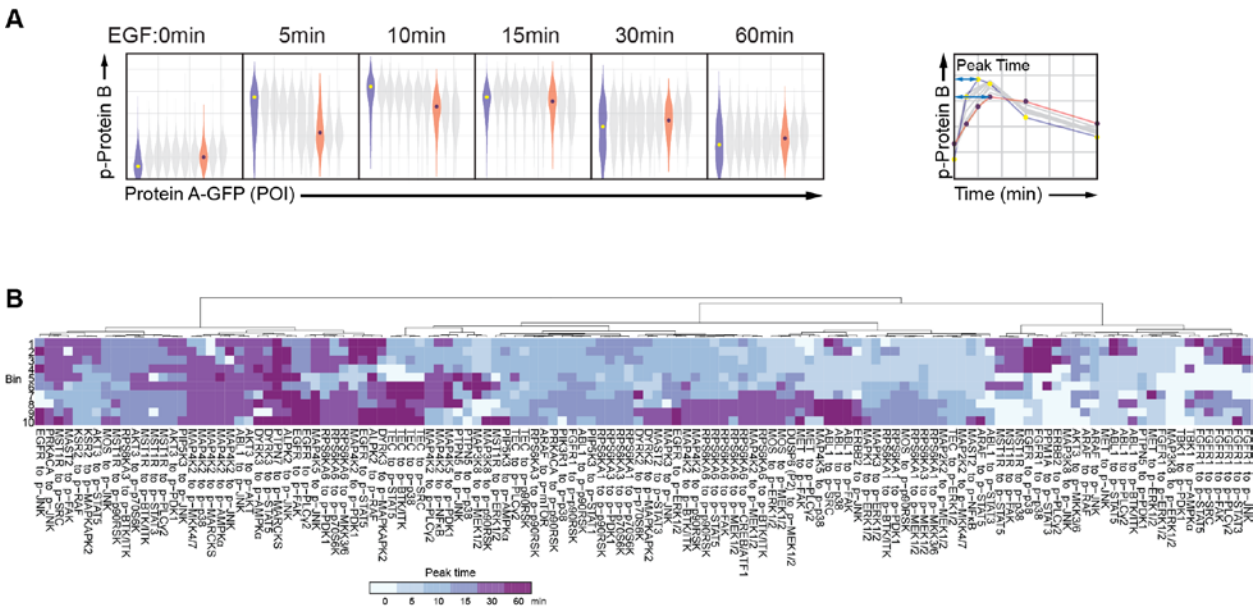



**Figure S6. Protein abundance-modulated signaling peak times. A**, Schematic plots of peak time analysis. The abundance levels of overexpressed GFP-POI were split into ten bins (left). Median phosphorylation in each bin over the one-hour EGF stimulation time course are plotted to visualize abundance dependency of signaling peak times (right). **B**, Heat map shows pairs of signaling relationships with strong protein abundance-influenced signaling peak times.

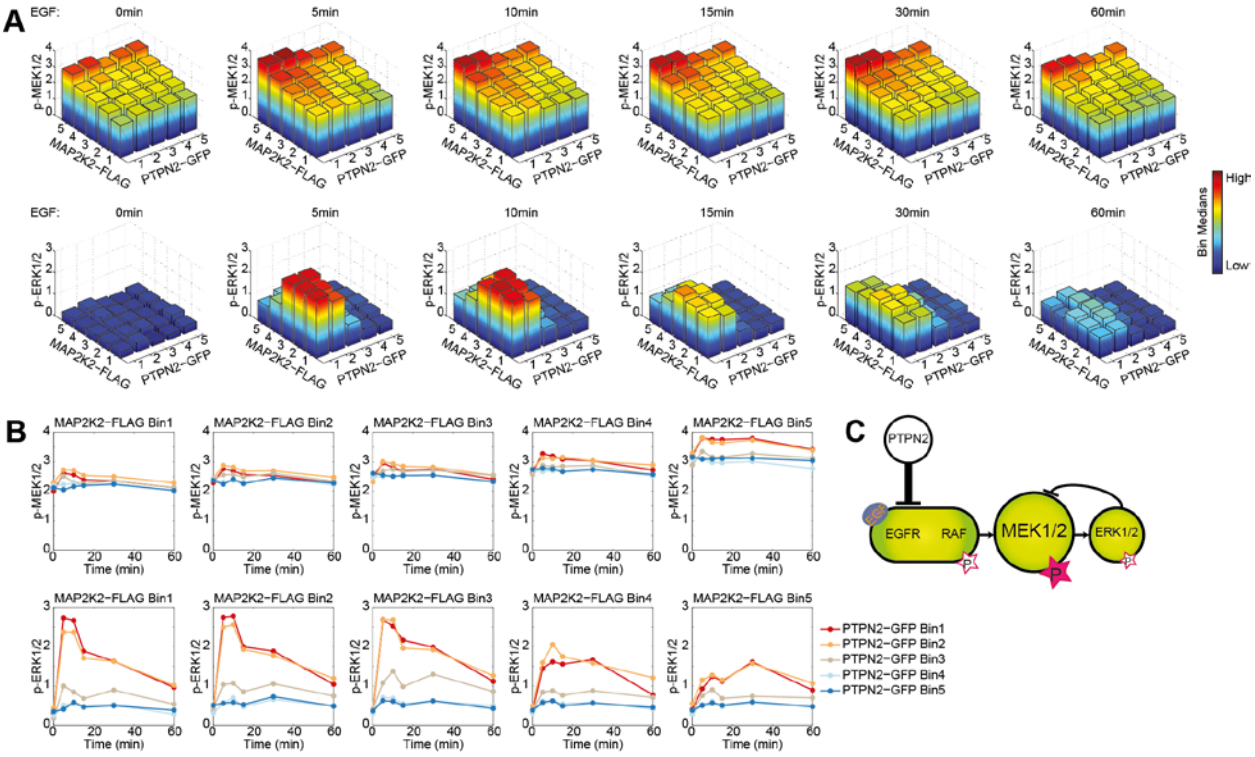



**Figure S7. Kinase and phosphatase pairwise overexpression.** **A**, For the pairwise overexpression of MAP2K2-FLAG and PTPN2-GFP, median phosphorylation levels of p-MEK1/2 and p-ERK1/2 for all bins over the 1-hour EGF stimulation time course are shown with the bar heights and bar colors. **B**, Signaling trajectories of p-MEK1/2 and p-ERK1/2 are plotted by connecting the medians of each individual bin over the 1-hour EGF stimulation time course. **C**, Schematic illustrates the modulation of RAF/MEK/ERK cascade signaling states and dynamics by the pairwise overexpression.

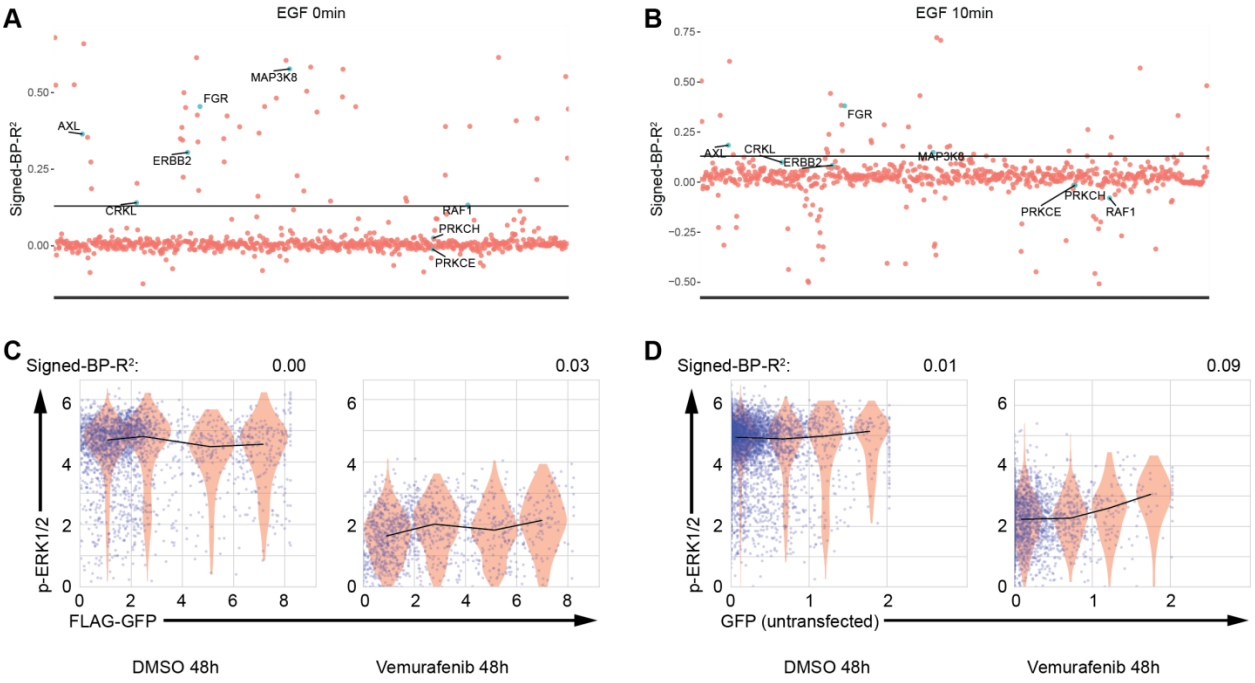



**Figure S8. Kinase overexpression induces vemurafenib resistance in the melanoma A375 cell. A-B**, Signed-BP-$R^2$ values of each overexpressed POI to p-ERK1/2 (**A**) before and (**B**) after 10-minute EGF stimulation. Kinases previously shown to induce RAF inhibitor resistance in A375 cells (identified in Johannessen et al., 2010) are labeled. **C-D**, Controls of FLAG-GFP overexpression and untransfected cells did not have strong signaling relationship to p-ERK1/2 after 48-hour incubation with vemurafenib or DMSO mock control.

# Result 3: CellCycleTRACER Accounts for Cell Cycle and Volume in Mass Cytometry Data

Maria Anna Rapsomaniki[1*], Xiao-Kang Lun[2,3*], Stefan Woerner[1], Marco Laumanns[1], Bernd Bodenmiller[2#] and María Rodríguez Martínez[1#]

[1] IBM, Zürich Research Lab, Säumerstrasse 4, 8803 Rüschlikon, Switzerland

[2] Institute of Molecular Life Sciences, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[3] Molecular Life Science Ph.D. Program, Life Science Zürich Graduate School, ETH Zürich and University of Zürich, Zürich, Switzerland

* Contributed equally

# Shared last authorship

Correspondence should be addressed to

B.B. (bernd.bodenmiller@imls.uzh.ch)

M.R.M. (mrm@zurich.ibm.com)

**Abstract**

Recent studies have shown that cell cycle and cell volume are confounding factors when studying biological phenomena in single cells. Here we present a combined experimental and computational method, CellCycleTRACER, to account for these factors in mass cytometry data. CellCycleTRACER is applied to mass cytometry data collected on three different cell types during a TNF-α stimulation time-course. CellCycleTRACER reveals signaling relationships and cell heterogeneity that were otherwise masked.

**Introduction**

Single-cell analysis technologies are rapidly improving and will soon match the performance of their population-level counterparts. RNA transcriptomes can be quantified in thousands of single cells, and analyses of transcriptomes of single cells with spatial resolution in tissues have been reported (Chen et al., 2015; Tang et al., 2009). Mass cytometry has the potential to enable simultaneous detection of up to 50 proteins and protein modifications, such as phosphorylation sites, in single cells (Bandura et al., 2009; Bodenmiller et al., 2012). Recent developments enable highly multiplexed imaging of similar numbers of markers in adherent cells and tissues (Bodenmiller, 2016; Giesen et al., 2014; Schapiro et al., 2017).

Single-cell data are typically used to identify cell subpopulations that share similar transcript or protein expression or functional markers. Analyses of these subpopulations can be used to reveal differences between tissue compartments in health and disease (Amir et al., 2013; Bruggner et al., 2014; Di Palma and Bodenmiller, 2014; Qiu et al., 2011), to reconstruct signaling network interactions, to study regulatory mechanisms (Krishnaswamy et al., 2014; Lun et al., 2017), and, together with clinical data, to identify single-cell features that predict characteristics such as response to treatment and likelihood of relapse (Chevrier et al., 2017b). For continuous processes, such as stem cell differentiation and the cell cycle, single-cell data allow the *in silico* reconstruction of the temporal dimension and thus the investigation of the underlying molecular changes and circuitries. Several algorithms designed to reconstruct cell trajectories from single-cell data are available, each with distinct strengths and weaknesses (Bendall et al., 2014; Gut et al., 2015; Kowalczyk et al., 2015; Setty et al., 2016).

Recent single-cell transcriptomic studies revealed that cell-cycle state and cell volume contribute to phenotypic and functional cell heterogeneity even in monoclonal cell lines (Buettner et al., 2015; Padovan-Merhar et al., 2015). This heterogeneity can obscure biological phenomena of interest (McDavid et al., 2016; Patel et al., 2014). For analysis of single-cell

transcriptomic data, computational methods have been developed to reveal variability in cell-cycle state and cell volume; these methods use principal component analysis, random forests, LASSO, logistic regression, support vector machines, and latent variable models (Buettner et al., 2015; McDavid et al., 2016; Scialdone et al., 2015; Wilson et al., 2015). These methods leverage large numbers of previously annotated cell-cycle genes and are thus not transferrable to mass cytometry data analyses.

Here, we developed a combined experimental and computational method, called CellCycleTRACER, to quantify and correct cell volume and cell cycle effects in mass cytometry data. The application of CellCycleTRACER to measurements of three different cell lines over a 1-h TNF-α stimulation time course revealed signaling features that had been otherwise confounded by cell cycle and cell volume effects.

## Results

### ***Cell cycle and volume effects measured by mass cytometry***

The impact of cell-cycle and cell-volume heterogeneity on mass cytometry data has not been addressed. We, therefore, set out to characterize how these factors influence commonly employed mass cytometry data analyses. To assess the effect of cell cycle, we exploited the simultaneous measurements of four cell-cycle markers recently identified by Behbehani et al., 2012: phosphorylated histone H3 (p-HH3), which peaks in the mitotic phase; phosphorylated retinoblastoma (p-RB), which monotonically increases from late G1 to M phase; cyclin B1, which increases from G2 to early M phase and rapidly diminishes during the late M phase; and 5-Iodo-2′-deoxyuridine (IdU), a thymidine analog incorporated during the S phase. We found that cell signaling as measured by protein phosphorylation strongly depended on the cell-cycle phase (Supplementary Note 1 and Supplementary Fig. 1). For example, a biaxial plot of phosphorylation of Ser241 on PDK1 versus phosphorylation of Thr172 on AMPKα revealed that in G2 and M phases, phosphorylation levels were elevated (Fig. 1a). Consequently, the estimated Pearson correlation coefficient between these two markers appears to be high due to the G2 and M cells that inflate the correlation. Less dramatic cell-cycle effects were also observed in published data (Behbehani et al., 2012) from a population of human T cells analyzed using a panel of immune-related cell-surface markers (Supplementary Fig. 2).

To assess the impact of cell volume, we had first to identify a marker that could be used to robustly quantify cell volume at a single-cell level. The ruthenium complex bis(2,2'-bipyridine)-

4’-methyl-4-carboxybipyridine-ruthenium-N-succidimyl ester-bis(hexafluorophosphate) (ASCQ_Ru) stains proteins by covalently binding to amino groups (Tokarski et al., 2006) (Supplementary Fig. 3a). ASCQ_Ru can be used in mass cytometry to reliably measure cell volume, as demonstrated using confocal laser scanning microscopy and three-dimensional cell reconstruction (Supplementary Fig. 3b-d), provided that the cells are not under conditions where the total protein mass and volume become uncorrelated (e.g., under drastic changes in osmolarity). In mass cytometry, ASCQ_Ru is measured by the ion counts of seven ruthenium isotopes ($^{96}$Ru, $^{98}$Ru, $^{99}$Ru, $^{100}$Ru, $^{101}$Ru, $^{102}$Ru, $^{104}$Ru) that do not overlap with channels used for antibody measurements. Similarly to what we found for the cell-cycle, the estimated correlation coefficients among phosphorylation markers were influenced by cell-volume heterogeneity (Fig. 1b, Supplementary Fig. 4). Given that the cell-cycle state and cell volume broadly confounded marker relationships, it was not surprising that analyses of mass cytometry data using standard statistical approaches, such as Pearson or Spearman correlations, or state-of-the-art computational methods, such as tSNE or DREMI, resulted in misleading conclusions (Supplementary Figs. 1, 4).

***CellCycleTRACER: Normalizing cell cycle and volume effects***

Cell volume and cell cycle change in a continuous manner and should be corrected, or at least taken into account, accordingly. Therefore, we developed CellCycleTRACER, an algorithm for the analysis of single-cell mass cytometry data that enables correction for cell-cycle-state and cell-volume heterogeneity. CellCycleTRACER is implemented as a simple and intuitive graphical user interface and can be applied to any mass cytometry dataset. Its application requires that four channels be dedicated to the cell-cycle markers p-HH3, p-RB, cyclin B1, and IdU (see Software and Data Availability).

CellCycleTRACER first exploits the ASCQ_Ru signal to transform raw marker counts into single-cell volume-relative intensities (Fig. 1c, Methods and Supplementary Note 2). After cell-volume correction, CellCycleTRACER uses data on the aforementioned four cell-cycle markers to classify cells into discrete cell-cycle phases and order them on a continuous path analogous to cell-cycle pseudotime (Fig. 1d, Methods and Supplementary Notes 3, 4). To automatically classify cells according to cell-cycle stage, CellCycleTRACER exploits a new machine learning approach that combines decision trees and Gaussian mixture models (Supplementary Figs. 5, 6, Methods and Supplementary Note 3); this approach reproduced manual gating procedures with 98.9% accuracy (Fig. 1d, Supplementary Fig. 7). Next, the single

cells are ordered on a continuum that traces cell-cycle evolution based on a novel trajectory reconstruction technique (Fig. 1d, e). To achieve this, CellCycleTRACER exploits the prior cell-cycle phase assignment and identifies the optimal one-dimensional embedding of the four cell-cycle markers that preserves the known order of the cell-cycle phases by minimizing ordering violations (Fig. 1e, Supplementary Fig. 8, Methods and Supplementary Note 4). Finally, the cell-cycle trajectories of the measured markers are obtained by projecting single-cell measurements onto the pseudotime dimension (Fig. 1d, lower right).

Reconstructed cell-cycle trajectories of the four markers used for the pseudotime inference (p-HH3, p-RB, cyclin B1, and IdU) agreed with their cell cycle-dependent variation (Fig. 1d, dashed lines)(Behbehani, 2018; Behbehani et al., 2012). Additionally, the pseudotime was validated by analyses of two independent cell cycle markers, p-CDK1 (Tyr15) and cyclin E (Fig. 1d, solid lines). CellCycleTRACER results faithfully recapitulated prior biological knowledge. Phosphorylation of Tyr15 on CDK1 progressively increased during S and G2 phase, peaked at the G2/M transition and sharply decreased after the entry to M phase. Cyclin E progressively accumulated during the G1 phase and reached the maximum at the G1/S transition before being degraded during the S phase (Bertoli et al., 2013; Castedo et al., 2002) (Fig. 1d, Supplementary Fig. 9). Furthermore, comparison with five state-of-the-art trajectory and embedding reconstruction methods showed that these methods failed to reproduce biologically relevant orderings of the cell cycle (Supplementary Fig. 10). For example, Wanderlust ordered the cells from G1→G2→S→M. Since Wanderlust works by first constructing *k* I-nearest neighbor graphs in the four-dimensional space of the cell-cycle markers and assumes that changes in protein abundance levels are gradual in the trajectory, it traversed the data in the wrong order because the G1 cluster is closer to G2 than S due to the jump in IdU. The other methods tested resulted in different incorrect orderings. SCUBA constructed a G2→S→G0/G1 trajectory and incorporated the M phase cells in the other clusters; TSCAN constructed a M→S/G2→G0/G1 trajectory by mixing together G2 and S cells; and Monocle ordered the data as G0/G1→M→G0/G1→G2→S, by ordering M phase cells in the middle of the G0/G1 cluster. Last, diffusion maps yielded a non-linear, low-dimensional embedding of the data that did not capture the known ordering. Since these methods are unsupervised techniques, they reconstruct continuous trajectories of the given measurements with no additional label information. It is thus impossible to “force” these methods traverse the data in the known cell-cycle phase order. CellCycleTRACER, however, exploits the known order of the phases through a mathematically well-defined optimization routine and guarantees by design that the known ordering will be preserved in the inferred one-dimensional embedding.

Reconstructed cell-cycle trajectories of cell-surface markers from a population of human T cells (Behbehani et al., 2012) indicated a continuous increase across the cell cycle for many of the proteins, peaking at the M phase (Supplementary Fig. 11). CellCycleTRACER can also remove cell-cycle-related inter-sample variations (due, for example, to the use of different cell lines or of the same cell line at different stimulation time points) and enables unbiased multi-sample analyses by trajectory alignment. This is achieved using a subsampling strategy that equalizes the relative cell-cycle phase proportions either to the mean inter-sample proportions or to the proportions of a user-selected sample (Fig. 1f, Methods and Supplementary Note 5). Last, CellCycleTRACER can correct for cell-cycle-related intra-sample variations by dividing the ordered single-cell values by the normalized mean trajectory (Fig. 1g, Methods and Supplementary Note 6). The dataset can be exported after any step of the pipeline, facilitating the use of various downstream data analysis approaches.

***Assessing CellCycleTRACER with TNF-α stimulation data***

To test the performance of our method, we measured the abundances of 25 protein phosphorylation sites, three housekeeping proteins, and three phenotypical markers in conjunction with the cell-volume and cell-cycle markers (Supplementary Table 1) in HEK293T (embryonic kidney), MDA-MB-231 (breast cancer), and THP-1 (monocyte) cells that had been stimulated with TNFα for 0, 5, 10, 15, 30, and 60 minutes (Methods). Analyses of cell volume at the control time point (0 minute) showed that MDA-MB-231 cells had on average the largest volume, followed by HEK293T and THP-1 cells (Fig. 2a). After cell-volume correction using CellCycleTRACER, the single-cell volume distributions in the three cell lines perfectly aligned (Fig. 2a). Marker abundances were strongly influenced by the cell-volume correction. For example, the amount of phosphorylated MKK4 (Ser257/Thr261) in THP-1 cells was 2-fold lower compared to the amounts in the other two cell lines when uncorrected for volume biases; after the correction, the amounts were nearly identical in each of the cell lines (Fig. 2a). After cell-volume correction, the coefficients of variation of the measured markers were reduced, indicating that our method corrected for cell-volume-dependent variations (Fig. 2a, bottom).

We next analyzed the cell-cycle evolution of different phosphorylation markers in response to TNFα stimulation. This analysis exposed cell-cycle-specific phosphorylation responses to stimulation. For example, in THP-1 cells, phosphorylation of p38 (Thr180/Tyr182) in response to TNFα stimulation was 2-fold stronger in G2/M phases compared to G0/G1 phase (Fig. 2b, left). The cell-cycle dependency of p38 phosphorylation was confirmed by flow

cytometry analysis where a similar fold change across the cell cycle phases was observed (Supplementary Fig. 12). It was reported previously that TNFα induces histone H3 phosphorylation that peaks at 30 minutes post-stimulation; this contributes to chromatin remodeling and enhances accessibility of DNA to transcriptional factor NFκB(Seidel et al., 2011). Analysis of the TNFα-stimulated THP-1 data using CellCycleTRACER revealed that this effect was cell-cycle dependent, as levels of phosphorylated histone H3 (Ser28) in the build up to the S phase were twice as high as in early G0/G1 or G2 phases of the cell cycle (Fig. 2b, right). The application of CellCycleTRACER aligned the trajectories and removed the bias introduced by the cell-cycle stage (Supplementary Fig. 13).

We next assessed the performance of CellCycleTRACER by comparing data before and after cell-volume and cell-cycle correction. First, Pearson correlation, Spearman correlation, and DREMI (a mutual information based metric) (Krishnaswamy et al., 2014) were used to quantify the relationship strength between measured markers in the unstimulated THP-1 cell data. As expected, after cell-volume and cell-cycle correction, Pearson correlation, Spearman correlation, and DREMI values for two cell-volume markers, ASCQ_$^{102}$Ru and ASCQ_$^{104}$Ru decreased significantly (Fig. 2c). For the signaling relationship between p-PDK1 and p-AMPKα, which is also affected by cell-cycle stage (Fig. 1a), Pearson correlation, Spearman correlation, and DREMI values were reduced from 0.88, 0.58, and 0.60 to 0.49, 0.34, and 0.34, respectively, upon application of CellCycleTRACER (Fig. 2c). Importantly, CellCycleTRACER correction had a smaller effect on the known direct signaling relationship of p-ERK (Thr202/p-Tyr-204) to p-p90RSK (Ser380), with a Pearson correlation slightly reduced from 0.77 to 0.69, Spearman correlation reduced from 0.75 to 0.67, and DREMI value reduced from 0.55 to 0.45 upon CellCycleTRACER application indicating that our method preserves real signaling relationships (Fig. 2c).

Second, we quantified the extent of cell-cycle-induced bias removed by CellCycleTRACER using an approach based on principal component analysis on a mixture of unstimulated (t=0) and stimulated (t=15 min) THP-1 cells (Fig. 2d). Specifically, after estimating the principal components of the data before and after cell-cycle correction, we fitted a linear model of the principal components on the cell-cycle-state index (i.e., G1, S, G2, and M phase) and the stimulation state and computed the variance explained by the fit ($R^2$) in all cases. Before correction, a large percentage of the variance in the first principal component was explained by the cell-cycle state; the effect was virtually eliminated by cell-cycle correction using CellCycleTRACER (Fig. 2d, left). Conversely, when the cell-cycle effect was eliminated, the

increase of $R^2$ for components 2 and 3 indicates that a larger percentage of the variance in the data was explained by the stimulation.

Third, we used CellCycleTRACER to assess the impact of correction on signaling network reconstruction with DREMI. Pairwise DREMI analysis on unstimulated THP-1 cells indicated that before cell-volume and cell-cycle corrections, phosphorylation sites known to be elevated in the M phase, such as Ser529 on NFκB, Thr172 on AMPKα, Thr334 on MAPKAPK2, and Ser241 on PDK1, were clustered together (Fig. 2e, left). After CellCycleTRACER was used to correct for heterogeneity in cell volume and cell cycle, DREMI scores were reduced in general, but a clear pattern consistent with MAPK/ERK and the AKT pathway activation appeared (Fig. 2e, right). Without correction for cell-volume and cell-cycle effects, signaling networks reconstructed with the top 10 signaling relationships as identified with DREMI in unstimulated THP-1 cells did not agree completely with commonly accepted signaling knowledge (Fig. 2f, left), whereas with the correction canonical relationships were seen (Fig. 2f, right) (Cohen and Frame, 2001; Manning and Cantley, 2007; Mendoza et al., 2011). Thus, pre-processing to correct mass cytometry data for cell-volume and cell-cycle heterogeneity is necessary for accurate analyses of correlation and variance, mutual information-based signaling relationship analysis (performed here with DREMI), and signaling network reconstruction.

Finally, we assessed how the cell-volume and cell-cycle corrections performed with CellCycleTRACER influenced the dimensionality reduction of a heterogeneous population of single cells with tSNE. Before the correction, the cell cycle confounded the separation of the cells in the tSNE plot (Fig. 2g, left), obscuring the cell line identities of the individual cells. The M phase cells from all analyzed cell lines were clustered, whereas cells from all other cell-cycle phases were mixed, and MDA-MB-231 cells were separated into three clusters. After cell-cycle correction, three clusters corresponding to the three different cell lines were observed in the tSNE plot, and the cell-cycle origin of each cell in each cluster appeared random (Fig. 2g, right).

## Discussion

In summary, cell volume and cell cycle can confound downstream mass cytometry data analysis. The presented experimental and computational approach, which we call CellCycleTRACER, corrects for the influences of volume and cell-cycle phase on mass cytometry data. CellCycleTRACER is a supervised manifold learning method that, in contrast to existing methods, exploits the cell-cycle phase labels to guarantee that the known ordering will be preserved in inferred embedding. With CellCycleTRACER we provide the mass cytometry

community with a method for supervised comprehensive analysis of cellular transitions. We expect that use of CellCycleTRACER will be particularly important when highly heterogeneous cell populations with deregulated cellular processes, as typically found in tumors, are analyzed.

**Acknowledgements**

We would like to thank the Bodenmiller laboratory for support and fruitful discussions, the Pelkmans laboratory for sharing experimental materials, Dr. Stéphane Chevrier for the help with flow cytometry analysis, Dr. Vinko Tosevski and Dr. Tess Brodie at the Mass Cytometry Facility, University of Zürich, and Dr. José María Mateos Melero at the Center for Microscopy and Image Analysis, University of Zürich for support and troubleshooting help. B.B.'s research is funded by an SNSF R`Equip grant, a SNSF Assistant Professorship grant (PP00P3-144874), by the European Research Council (ERC) under the European Union's Seventh Framework Program (FP/2007-2013)/ERC Grant Agreement n. 336921 and an NIH grant (UC4 DK108132). B.B. and M.R.M. are also funded by the SystemsX MetastasiX grant.

**Author contributions**

X.–K.L. and B.B. conceived the study and experiments. X.-K.L. developed reagents and performed all experiments. X.-K.L., M.R., and M.R.M. performed data analysis. M.R. and M.R.M. conceived the CellCycleTRACER algorithm. X.-K.L. and B.B. performed the biological analysis and interpretation. S.W. and M.L. conceived the optimization algorithm. B.B., M.R., X.-K.L and M.R.M. wrote the manuscript with input from all authors.

**Competing financial interest**

The authors declare no competing financial interest.

**Figure captions**

***Figure 1*.** Cell-volume and cell-cycle biases in mass cytometry data and their corrections using CellCycleTRACER. **(a)** Biaxial plot of p-PDK1 (Ser241) versus p-AMPKα (Thr172) in THP-1 cells, where pre-gated cell-cycle phases are indicated by different colors. Computation of Pearson correlation coefficients across cell-cycle phases indicates a strong cell-cycle bias. **(b)**

Biaxial plot of p-PDK1 (Ser241) versus p-AMPKα (Thr172) in G0/G1 phase THP-1 cells that were pre-gated by cell volume as indicated by different colors. Pearson correlation coefficients are indicative of the cell-volume bias. **(c)** Cell-volume correction using ASCQ_Ru measurements removes cell-volume variability and transforms raw counts of measured markers into relative concentrations at single-cell resolution. **(d)** Construction of cell-cycle pseudotime initiates with automatic classification of the cells into discrete cell-cycle phases using measurements of IdU, cyclin B1, p-HH3, and p-RB (Behbehani et al., 2012). The optimal trajectory across phases is constructed by projecting the data in a one-dimensional embedding function analogous to cell-cycle pseudotime. Mean trajectories of all measured cell-cycle markers across the reconstructed pseudotime recapitulate known behavior. Markers used to construct the pseudotime (IdU, cyclin B1, p-HH3, and p-RB) are shown as dashed lines, additional cell cycle markers used as validation (cyclin E and p-CDK1) are shown as solid lines. **(e)** Simplified example of the trajectory reconstruction technique. By exploiting prior information of the class labels for each cell and the order of the classes, the best embedding function is computed by selecting the one that optimally preserves the known ordering across all cells in the new subspace defined by the embedding. **(f)** CellCycleTRACER aligns cell-cycle pseudotime by equalizing cell-cycle phase duration across all analyzed samples. **(g)** CellCycleTRACER correction for cell-cycle redistributes the single cells independently of cell-cycle variation.

***Figure 2.*** CellCycleTRACER corrects for cell-volume and cell-cycle heterogeneity enabling unbiased data visualization and downstream analysis. **(a)** Overlaid histograms reveal differential data observations before and after cell-volume correction. Bar charts show that cell-volume correction also reduces intra-sample variation as coefficients of variation of measured markers decrease. **(b)** Abundance of p-p38 (Thr180/Tyr182) and p-HH3 (Ser28) plotted on the cell-cycle pseudotime based on data from TNFα-stimulated THP-1 cells. Stimulation time points are indicated by different colors. **(c)** Biaxial plots show signaling relationships between measured markers before and after cell-volume and cell-cycle correction. Relationship strengths quantified by Pearson correlation, Spearman correlation, and DREMI are indicated in the corresponding barplots. **(d)** Principal component analysis of data originating from a mixed population of unstimulated (t=0 min) and stimulated (t=15 min) THP-1 cells. After computing the principal components of the data before and after cell-cycle correction, the variances explained by fitting a linear model of the principal components on the cell-cycle state index (left) and the stimulation state (right) were estimated, indicating removal of cell-cycle confounding effects. **(e)**

Clustergrams of pairwise DREMI analyses of unstimulated THP-1 cells before and after cell-volume and cell-cycle corrections. After the removal of cell-volume and cell-cycle variability, DREMI scores of non-interactive pairs are reduced, and AKT and MAPK/ERK signaling pathways become apparent. **(f)** Network reconstruction using the top 10 DREMI scorers in unstimulated THP-1 cells before and after cell-volume and cell-cycle corrections. Network reconstructed after correction recapitulates key regulatory interactions in the AKT and MAPK/ERK pathways. **(g)** tSNE maps of THP-1, MDA-MB-231, and HEK293T cell lines before and after cell-volume and cell-cycle correction. Cell-cycle and cell-volume markers were not included in the tSNE analysis.

## Methods

### ***Cell culture***

HEK293T, MDA-MB-231, and THP-1 cells were obtained from ATCC. HEK293T cells were cultured in DMEM (D5671, Sigma) supplemented with 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin. MDA-MB-231 and THP-1 cells were cultured in Leibovitz's L-15 Medium (11415064, Gibco) and RPMI-1640 Medium (52400025, Gibco), respectively, both supplemented with 10% FBS, 100 U/ml penicillin, and 100 µg/ml streptomycin. For passaging or harvesting, HEK293T and MDA-MB-231 cells were first detached by incubating with 1X TrypLE™ Express (Life Technologies) for 2 minutes at 37 °C.

### ***TNFα stimulation***

HEK293T, MDA-MB-231, and THP-1 cells were seeded in 6-well plates at densities of 0.7 million cells, 0.5 million cells, and 1 million cells per well, respectively. After 2 days, cells were stimulated with TNFα (R&D Systems) at 10 ng/ml. Aliquots were collected for analysis at 0, 5, 10, 15, 30, and 60 minutes (stimulation was performed in reverse order to enable simultaneous harvesting of all conditions). At 20 minutes before harvesting, 5-iodo-deoxycytidine (IdU) was added to the medium at the final concentration of 10 µM. At 2 minutes before harvesting HEK293T and MDA-MB-231 media were replaced with 1X TrypLE to induce detachment. At the time of harvest, paraformaldehyde (PFA, Electron Microscopy Sciences) was added to the cell suspension at a final percentage of 1.6%, and samples were incubated at room temperature for 10 minutes. Crosslinked cells were washed twice with cell staining media (CSM, PBS with 0.5% BSA, 0.02% $NaN_3$). After removal of supernatant, ice-cold methanol was used to resuspend the cells, followed by a 10-minute permeabilization on ice or long-term storage at -80 °C.

### ***Immunofluorescence and three-dimensional reconstruction***

CultureWell™ chambered coverglass wells (16-well, Thermo Fisher Scientific) were pre-coated by incubation with 10 µg/ml bovine plasma fibronectin (Thermo Fisher Scientific) at 37 °C for 1 hour. MDA-MB-231 cells were then seeded at a density of 1500 cells per well. On the second day, 4% PFA was used to crosslink the cells at room temperature for 20 minutes. The slide was then washed with PBST (0.5% Tween 20 in PBS) three times, and cells were subsequently permeabilized for 5 minutes with 0.1% TritonX-100 (diluted in PBS) at room temperature. After

washing with PBST three times, cells were incubated in blocking buffer (10% goat serum diluted in PBST) for 30 minutes at room temperature. Primary antibody (anti-GAPDH, 6C5, Thermo Fisher Scientific, 1 µg/ml; anti-Rab7, D95F2, Cell Signaling Technology, 1:100; or anti-β-actin, D6A8, Cell Signaling Technology, 1:200) was added, and slides were incubated overnight at 4 °C. Secondary antibody (goat anti-mouse Alexa Fluor® 488, 1:200 or goat anti-rabbit Alexa Fluor® 555, 1:200), supplemented with Hoechst 33342 at a final concentration of 100 µg/ml, was applied, and slides were incubated for 1 hour at room temperature. Antibodies were diluted in the blocking buffer. Slides were washed three times with PBST after each incubation step. For cell-volume analyses, cells were stained for total proteins with Alexa Fluor® 647 NHS ester (Thermo Fisher Scientific) for 10 minutes at a final concentration of 1 µg/ml. Slides were then mounted with ProLong Gold Antifade Reagent (Life Technologies) before imaging with a CLSM Leica SP5 microscope. Stacks were imaged every 0.5 µm, and the three-dimensional reconstruction and quantification of the total cell volume was performed with Imaris 7.7.2.

***Antibody conjugation***

Isotope-labeled antibodies were generated with MaxPAR antibody conjugation kit (Fluidigm) using the manufacturer's standard protocol. The antibody yield was determined based on absorbance of 280 nm. Candor PBS antibody stabilization solution (Candor Bioscience GmbH) was used to dilute antibodies for long-term storage at 4 °C.

***Barcoding and staining protocol***

Formalin-crosslinked and methanol-permeabilized cells were washed three times with CSM and once with PBS. Cells were incubated in PBS containing barcoding reagents ($^{105}$Pd, $^{106}$Pd, $^{108}$Pd, $^{110}$Pd, $^{113}$In, $^{115}$In, and $^{139}$La) at a final concentration of 50 nM for 30 minutes at room temperature and then were washed three times with CSM (Lun et al., 2017). Barcoded cells were pooled and stained with the metal-conjugated antibody mix (Supplementary Table 1) at room temperature for 1 hour. The antibody mix was removed by washing cells three times with CSM and once with PBS. For DNA staining, iridium-containing intercalator (Fluidigm) was diluted in PBS with 1.6% PFA and incubated with the cells at 4 °C overnight. On the day before measurement, the intercalator solution was removed, and cells were washed with CSM, PBS, and doubly distilled $H_2O$ sequentially. Total protein staining was performed with 25 µg/ml ASCQ_Ru (96631, Sigma) in 0.1 M $NaHCO_3$ solution for 10 minutes at room temperature. Cells

were then washed with CSM, PBS, and doubly distilled $H_2O$ sequentially. After the last wash, cells were resuspended in doubly distilled $H_2O$ and filtered through a 70-µm strainer.

***Mass cytometry analysis***

EQ™ Four Element Calibration Beads (Fluidigm) were added to the cell suspension at a 1:10 ratio (v/v). Samples were analyzed on a Helios mass cytometer (Fluidigm). The manufacturer's standard operation procedures were used for acquisition at a rate of ~200 cells per second. After data acquisition, all .fcs files from the same barcoded sample were concatenated. Data were then normalized, and bead events were removed (Finck et al., 2013). Doublets were removed, and cells were de-barcoded into their corresponding wells using a doublet-filtering scheme and single-cell deconvolution algorithm (Zunder et al., 2015). Subsequently, data were processed using Cytobank (http://www.cytobank.org/). Additional gating on the DNA channels ($^{191}$Ir and $^{193}$Ir) was used to remove remained doublets, debris, and contaminating particulates. Manual gating was performed on IdU, cyclin B1, p-HH3, and p-RB to identify cell-cycle stages (Behbehani et al., 2012).

***CellCycleTRACER workflow***

CellCycleTRACER requires as an input measurements of the four cell cycle markers (namely p-HH3, p-RB, cyclin B1, and IdU) as well as measurements of cell volume, ideally using the ASCQ_Ru markers.

***Data processing and cell-volume correction***

To determine cell volume at a single cell level, we initially experimented with three housekeeping proteins, namely GAPDH, actin, and RAB7 in HEK293T, MDA-MB-231, and THP-1 cells. Although all three proteins highly correlate with the total cell volume (Supplementary Fig. 14), single-cell measurements from a mixed population of three different cell lines revealed a large degree of cell-line-specific variability that surprisingly involved these housekeeping proteins (Supplementary Fig. 15). This indicated that these housekeeping proteins cannot be used for cell volume correction when heterogeneous populations are analyzed.

ASCQ_Ru (Supplementary Fig. 3a) is conventionally used in electrophoresis for the determination of protein abundance and has been reported to outperform other staining reagents with high sensitivity and large linear dynamic range of protein binding. Taking advantage of its additional fluorescent property, we validated ASCQ_Ru as a precise cell volume indicator using three-dimensional reconstruction of confocal images (Supplementary Fig. 3b-d)). Measurements across the three cell lines showed reduced cell line variability in comparison to housekeeping protein measurements (Supplementary Fig. 16).

CellCycleTRACER corrects the data (uploaded as .fcs files) on ASCQ_Ru to enable correction for cell-volume heterogeneity. Let $j = 1, \dots, m$ be the quantified protein marker in the $i = 1, \dots, n$ single cell. Let $y_{i,j}$ denote the abundance of marker $j$ in cell $i$ in raw experimental data. At the same time, let $v = 1, \dots, l$ denote the subset of the protein markers ($\{v\} \subset \{\mathrm{j}\}$) that contain information on the total cell volume – in our case the ASCQ_Ru markers. During the cell-volume correction step, CellCycleTRACER first normalizes the raw cell-volume measurements $y_{i,v}$ by dividing each marker by its mean value:

$$y_{i,v}^{norm} = \frac{y_{i,v}}{\frac{1}{n}\sum_{i=1}^{n} y_{i,v}}$$

The raw measurements of all $j = 1, \dots, m$ markers are then corrected for cell-volume variations by dividing $y_{i,j}$ by the mean value of $y_{i,v}^{norm}$:

$$y_{i,j}^{corr} = \frac{y_{i,j}}{\frac{1}{l}\sum_{v=1}^{l} y_{i,v}^{norm}} \quad (1)$$

Results of this process are shown in Supplementary Fig. 17. Comparison of measurements of the phosphorylated versus total amount of proteins MEK1/2 and ERK1/2 before and after cell volume correction indicated that the correction process was equally effective for both activated and total amounts of proteins (Supplementary Fig. 18). To avoid dividing by zero, CellCycleTRACER checks the data for zero values and, if found, substitutes zeros with the respective mean value. For more details on volume correction see Supplementary Note 2.

After cell-volume correction, selected channels of the raw measurements are transformed using the inverse hyperbolic sine function (asinh):

$$y_{i,j}^{trans} = \operatorname{asinh}\left(y_{i,j}^{corr}\right) = \ln\left(\frac{y_{i,j}^{corr}}{c} + \sqrt{\left(\frac{y_{i,j}^{corr}}{c}\right)^2 + 1}\right) \quad (2)$$

where the constant $c$, commonly referred to as the cofactor, is set to 5 according to the CyTOF community's standard practice. Unlike the standard logarithmic function that is undetermined at

zero values, asinh is linear around zero and becomes logarithmic beyond a threshold determined by the cofactor value. The overall effect of this transformation is to selectively compress large values, eliminating the typical long tails found in raw cytometry measurements. This results in a more symmetrical distribution that facilitates clustering and other machine learning analyses.

***Cell-cycle phase prediction***

After the volume correction, CellCycleTRACER classifies the single cells into discrete cell-cycle phases. To achieve this, we exploit the measurements of the four above-mentioned cell-cycle markers (IdU, p-HH3, cyclin B1, and p-RB) that are typically used in mass cytometry for manual cell-cycle gating. To eliminate possible biases introduced by variations in antibody concentrations and affinities, the data are standardized and set to have zero mean and unit variance. The prediction process is based on a hybrid approach that consists of two steps: (i) a classification step in which single-cell measurements of the four cell-cycle markers are given as input into a decision tree classifier to automatically predict the cell-cycle phase and (ii) a clustering step in which a Gaussian mixture model (GMM) is used to fit the data into clusters that represent the cell-cycle phases. The GMM is initialized using the predictions of the decision tree as prior knowledge.

Detailed descriptions of the implementation and performance on different datasets are given in Supplementary Note 3. In brief, we first used the four cell-cycle marker measurements from an experiment using THP-1 cells together with their class labels derived by manual gating (Supplementary Fig. 5a) to train a decision tree classifier. The resulting decision tree and the class proportions at the terminal nodes are shown in Supplementary Fig. 5b. We observed four pure terminal nodes, equal to the number of classes, indicating 100% classification accuracy in the training set. The order of the splits was identical to the order in the classification performed manually, indicating that the model faithfully captures the manual gating process. The decision tree accuracy in the independent test set was also 100%, meaning that all cells were correctly classified. After classification performance was validated, new experimental measurements were given as inputs and were automatically classified. The results on a HEK293T test set are shown in Supplementary Fig. 7a.

Second, the measurements were clustered using a GMM where the number of components was set to four, equal to the number of cell-cycle phases in our model. The parameters of the GMMs (mean vectors, covariance matrices, and class proportions) were

initialized using the decision tree predictions and iteratively refined until convergence using an expectation-maximization (EM) algorithm (Supplementary Fig. 6). After convergence, posterior probabilities of each of the four GMM components were computed, and the single cells were assigned to the component with the maximal posterior (results in Supplementary Fig. 7b).

This hybrid approach combines the intrinsic interpretability of decision trees, which enable extraction of a set of comprehensive if-else rules from the training data, with the probabilistic capabilities of the GMM framework. Specifically, decision trees are ideal for partitioning of a space using training data as prior knowledge, but they lack the notion of distribution and suffer from rigid boundaries. Since mass cytometry data are produced from different cell lines and possibly on different days, these data exhibit inter-experimental variability that makes the algorithm prone to misclassifying the points at the tails of the distribution. GMMs, on the other hand, allow flexible treatment of outliers. Shortcomings of GMMs – and other unsupervised clustering methods – are an inability to match the clusters to the known labels and no guarantee of convergence to the optimal solution. These limitations are especially acute when the classes are significantly imbalanced, as in the case of cell-cycle fractions which differ by an order of magnitude (e.g., G0/G1: 40-60%; M phase: 3-5% of the total cell population). By combining decision tree and GMM approaches we benefit from the advantages of decision trees to provide an initial guess close to the optimal solution and of GMMs to allow for a probabilistic interpretation of the class assignments. This refinement translates into better assignments for outliers and captures the classification uncertainty of cells transitioning between phases, a subtlety that is entirely missed by the decision tree.

***Trajectory reconstruction, alignment, and correction***

Cells progress along the cell cycle in a continuous way, gradually transitioning across consecutive phases whose boundaries are not always clearly defined, and exhibiting intra-phase variability (e.g., cells at early S and late S are drastically different). To better represent these pseudo-temporal fluctuations, we devised a method that reconstructs trajectories of biological cell-cycle time (*pseudotime*) from a population of unsynchronized single cells, ordering them according to cell-cycle progression. The details of the reconstruction method are given in Supplementary Note 4.

We assume that $n$ single cells are classified in four cell-cycle phases. Let $y_i$ denote a four-dimensional vector of cell-cycle marker abundances in each cell. We seek to construct a

one-dimensional embedding function of the four-dimensional vectors $y_i$, denoted as $f_\alpha(y)$, that represents pseudotime. One possible choice is to define $f_\alpha(y)$ as a linear combination of $y_i$:

$$f_\alpha(y_i) = \sum_{j=1}^{4} \alpha_j y_{i,j} \,,$$

where the coefficients $\alpha_j$ take values in $\mathbb{R}^4_{\geq 0}$. Under this formulation, our problem reduces to identifying a vector of coefficients $\alpha = (\alpha_1\ \alpha_2\ \alpha_3\ \alpha_4)$ that optimally maps the cell-cycle marker measurements to pseudotime. Since the ordering of the discrete classes is known a priori (G1→S→G2→M), we follow an optimization process that aims to guarantee this ordering in the desired embedding by minimizing the difference across cells that belong to adjacent classes (see also Supplementary Fig. 8). More specifically, for all cells $i_p, i_q$ belonging to adjacent classes $p, q$, we estimate $\alpha$ such that $f_\alpha\left(y_{i_p}\right) < f_\alpha\left(y_{i_q}\right)$, a constraint that translates into preserving the ordering in the embedding.

Collapsing the four-dimensional measurements into a lower-dimensional space may not result in fully separated clusters, with the implication that the ordering constraints might not be satisfied for all cells. To tackle this problem, we introduced slack variables into all constraints; these non-negative variables represent a degree of violation of the ordering constraint. We then minimize over a weighted sum of all slack variables, which leads to a mathematically well posed linear programming (LP) problem. Even though degenerate LPs can have multiple equivalent optima (convex set of optimal solutions), due to the presence of extrinsic and intrinsic variability in CyTOF data this does not occur in practice. Thus, the LP results in a single, optimal ordering. Since the solution time of the resulting LP grows substantially with the number of cells (Supplementary Fig. 19a) and can thus be computationally intensive, we randomly picked a fixed percentage of cells from each class and computed an optimal set of weights. A numerical investigation (Supplementary Fig. 19b) indicated that the solution of the LP, which yields the values of parameters $\alpha$, is robust to the sampling of cells even when a small percentage of cells is considered.

Once the values of parameters $\alpha$ are estimated, CellCycleTRACER visualizes the results by ordering the single cells based on their pseudotime values, resulting in single-cell trajectories for each marker. Additionally, CellCycleTRACER computes the mean trajectory of each marker by applying a mean filter on the single-cell trajectory, where the value for each cell is replaced by the mean of the neighboring cells in a sliding window of fixed size. Since different samples (e.g., different cell lines) can exhibit strong variations in the relative duration of the cell-cycle phases, CellCycleTRACER permits multi-sample analysis by either aligning the relative

cell-cycle phase proportions across individual samples to the mean vector of cell-cycle phase proportions or, alternatively, aligning them to one sample (e.g., one cell line) of interest (Fig. 1f, Supplementary Note 5). To remove the effect of the cell cycle on the marker measurements, CellCycleTRACER exploits the abovementioned mean trajectory, rescales it around 1 by dividing by the mean abundance of the marker and then divides the single-cell trajectory by the rescaled mean. This step removes cell-cycle-specific fluctuations and redistributes the single cells independently of cell-cycle variation (Fig. 1g, Supplementary Note 6). After every step of the analysis (e.g., cell-cycle classification, correction, alignment), data can be exported as .fcs files for further analysis.

## ***Implementation***

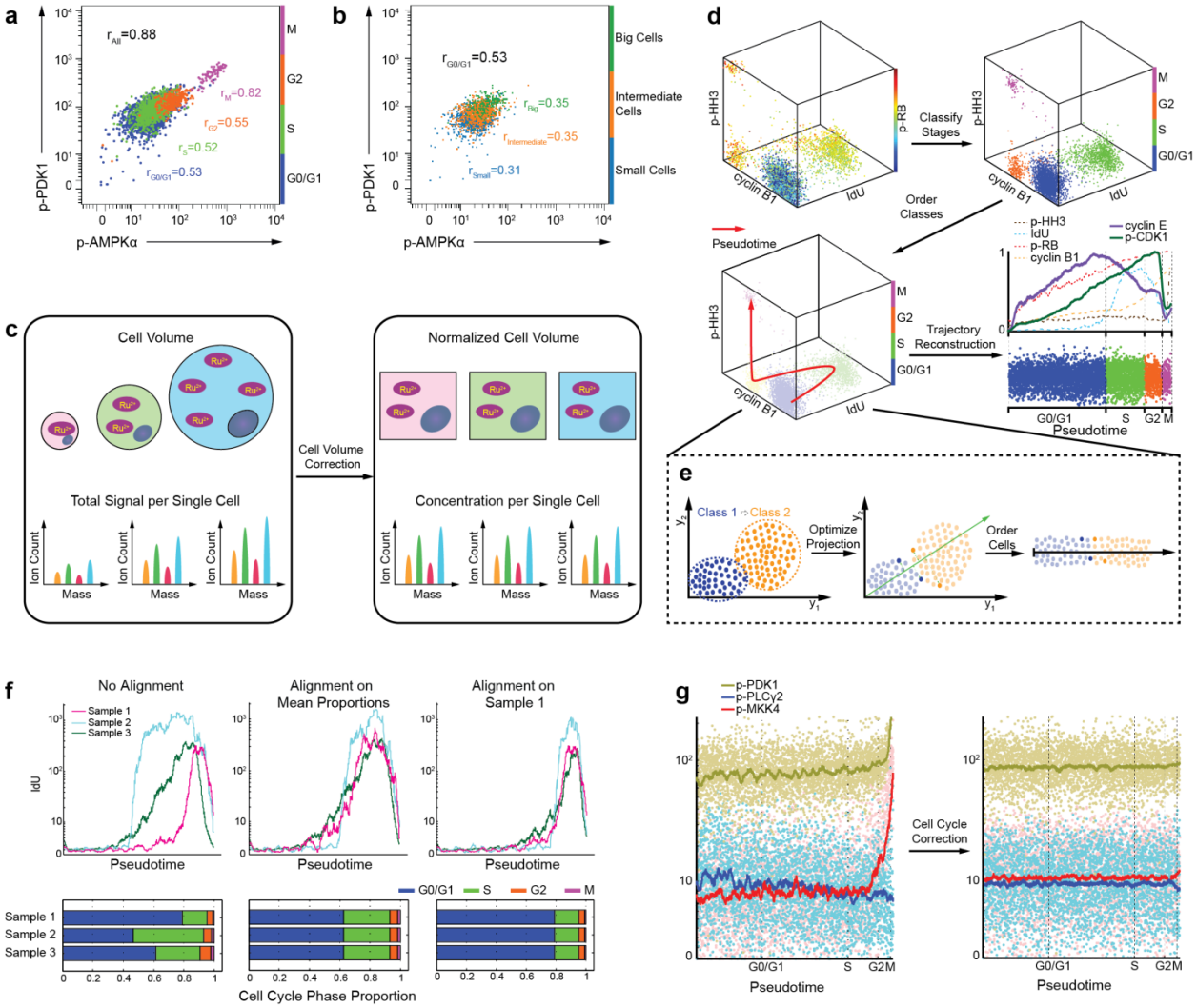
All methods were implemented using the Statistical and Optimization Toolboxes of MATLAB R2011b.

## ***Software and data availability***

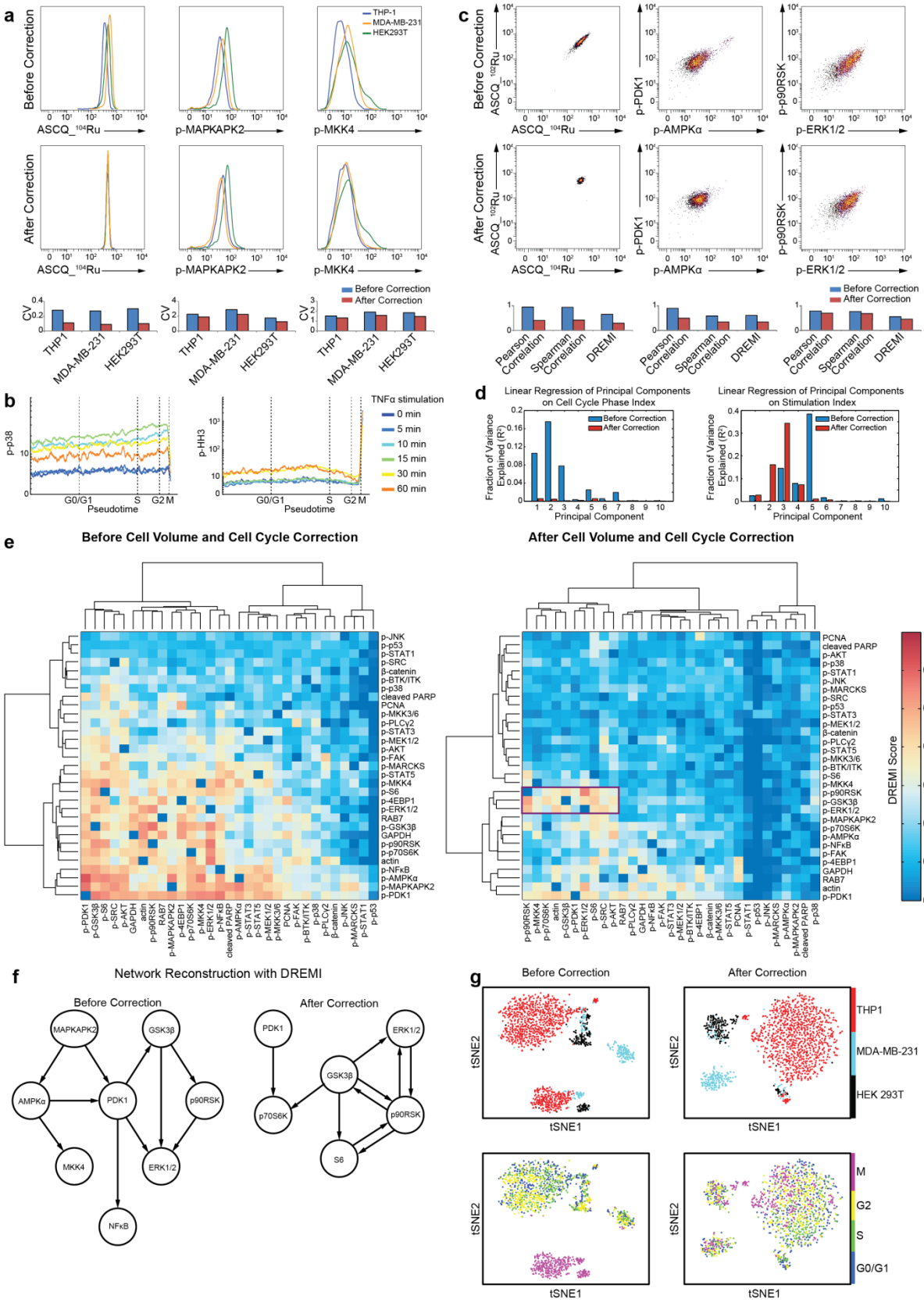
CellCycleTRACER is implemented as a web application, accessible using the following link: https://www.zurich.ibm.com/compsysbio/publications.html. CyTOF data for the three cell lines at all stimulation time-points are available on Cytobank under project 1129.

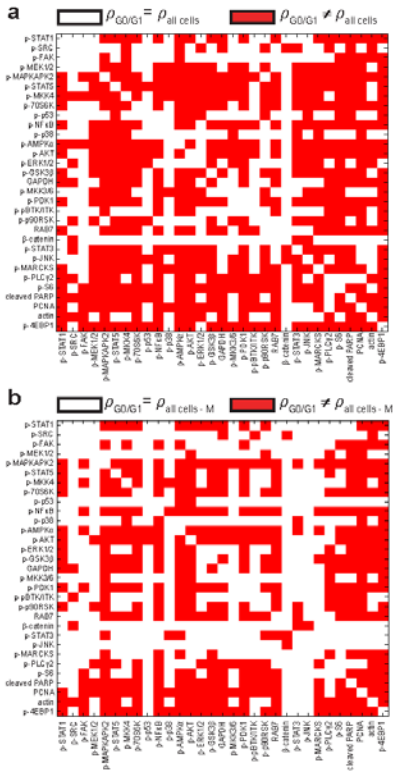
## Figures

Figure 1

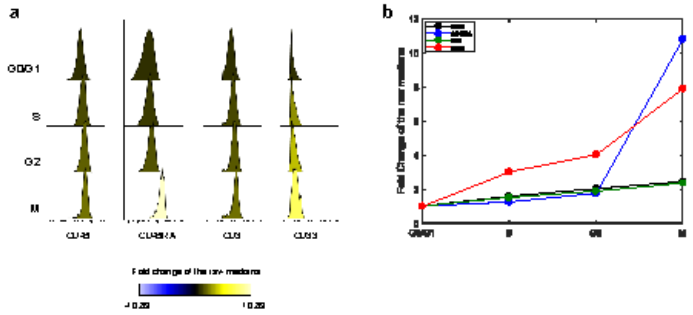


# Figure 2

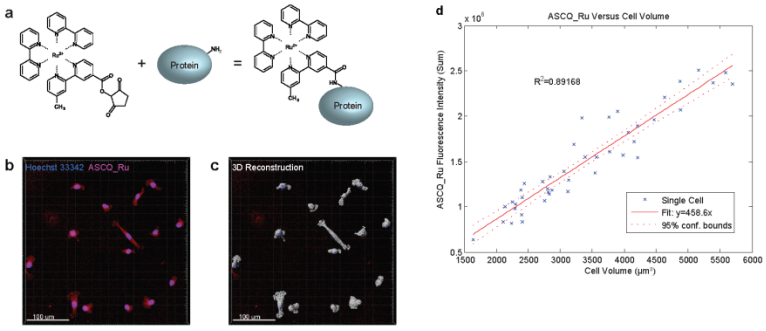

a
Before Correction
After Correction
ASCQ_104Ru
p-MAPKAPK2
p-MKK4
THP-1
MDA-MB-231
HEK293T
CV
Before Correction
After Correction
THP1
MDA-MB-231
HEK293T
b
p-p38
p-HH3
Pseudotime
G0/G1
S
G2 M
TNFα stimulation
0 min
5 min
10 min
15 min
30 min
60 min
c
Before Correction
After Correction
ASCQ_102Ru
ASCQ_104Ru
p-PDK1
p-AMPKα
p-p90RSK
p-ERK1/2
Pearson Correlation
Spearman Correlation
DREMI
d
Linear Regression of Principal Components on Cell Cycle Phase Index
Linear Regression of Principal Components on Stimulation Index
Fraction of Variance Explained (R²)
Principal Component
Before Correction
After Correction
e
Before Cell Volume and Cell Cycle Correction
After Cell Volume and Cell Cycle Correction
DREMI Score
f
Network Reconstruction with DREMI
Before Correction
After Correction
MAPKAPK2
GSK3β
AMPKα
PDK1
p90RSK
MKK4
ERK1/2
NFκB
p70S6K
S6
g
Before Correction
After Correction
tSNE1
tSNE2
THP1
MDA-MB-231
HEK 293T
M
G2
S
G0/G1



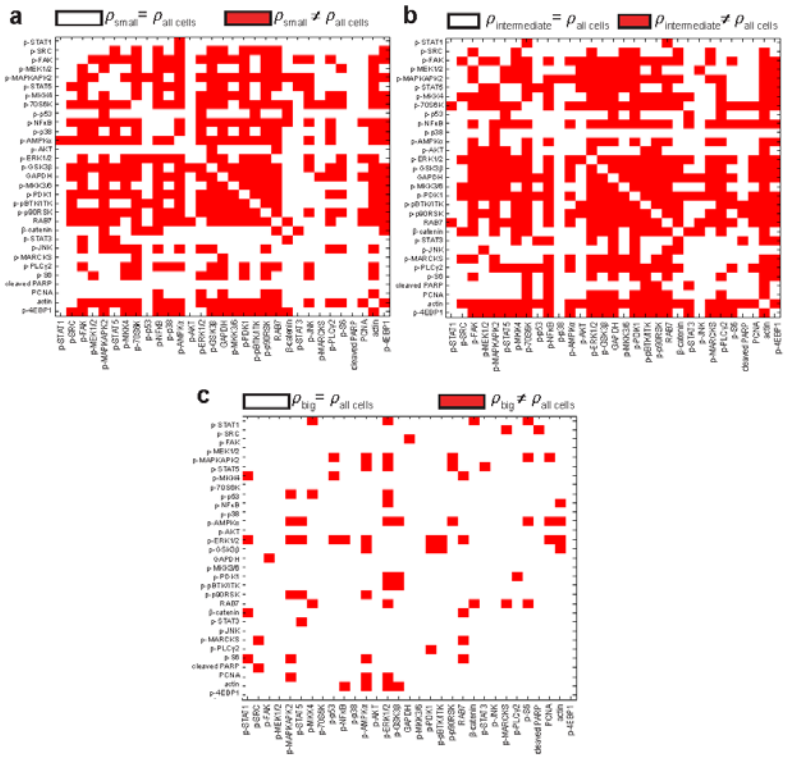

Supplementary Figure 1: **Pairwise correlation coefficients as computed from a mixed population of THP-1 cells are significantly influenced by the cell cycle state.** **(a)** Comparison of Spearman correlation coefficients computed using only G0/G1 cells, $\rho_{G0/G1}$, and using all cells, $\rho_{\text{all cells}}$ (details in Supplementary Note 1) indicates that the majority of protein pairs (292 out of total 465) exhibit statistically significant changes in correlation across the cell cycle (p-value $\leq 0.05$). **(b)** Excluding M phase cells from the data and repeating the analysis reveals a minor improvement in the introduced bias; however significant changes in the correlation coefficients are still observed in almost half of the protein pairs (222 out of 465).



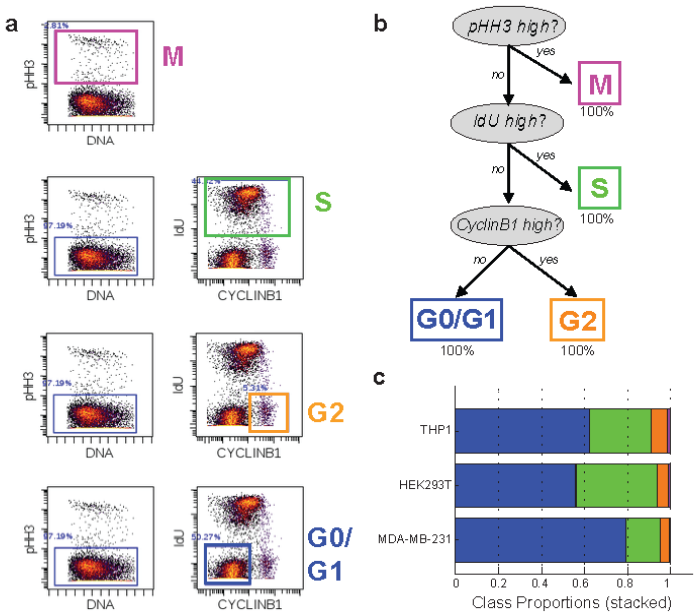

Supplementary Figure 2: **Cell cycle effects in cell surface markers.** To assess the effect of the cell cycle in a dataset that includes cell membrane markers, we analyzed the data from [1], where the authors used mass cytometry to analyze a population of human T cells and included in their panel cell surface markers (e.g., CD3, CD4, CD45) as well as cell cycle markers (e.g., p-HH3, p-RB, IdU, cyclin B1). When examining the distribution of the cell surface markers across the cell cycle phases, as identified by the authors via manual gating, we observed fluctuations in levels across the cell cycle for some of the measured markers **(a)**. Specifically, CD45, CD45RA, CD3, and CD33 progressively increase during the entire cell cycle, peaking at the M phase (2.5- to 11-fold increase with respect to G0/G1 phase **(b)**). Overall, this finding indicates that, even in well-studied systems where the cell cycle is not expected to be a prominent confounder, cell-cycle signatures can have a non-negligible imprint on the measured protein abundance.



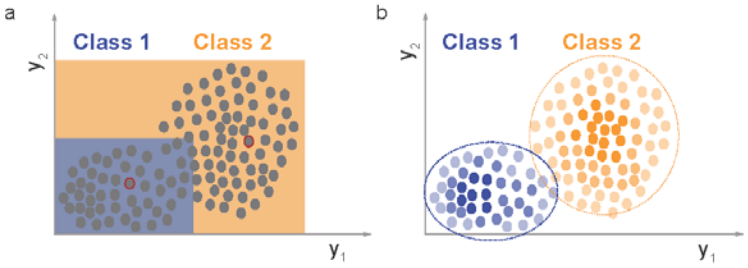

Supplementary Figure 3: **Validation of ASCQ_Ru as an indicator of cell volume.** **(a)** Structure of ASCQ_Ru and biochemical mechanism of ASCQ_Ru staining: ASCQ_Ru covalently binds to the amines on unspecific proteins. **(b)** Confocal images of MDA-MB-231 cells stained with ASCQ_Ru and Hoechst 33342 (DNA staining). **(c)** 3D reconstruction of imaged cells using ASCQ_Ru to determine cell volume. **(d)** Linear regression on summed ASCQ_Ru fluorescence intensity in single cells versus computed cell volume. Scale bar, 100$\mu$m.




Supplementary Figure 4: **Pairwise correlation coefficients are significantly influenced by cell volume**. Pairwise correlations computed using the whole cell population are significantly different (p-value $\leq 0.05$) than pairwise correlations computed using only **(a)** small cells (214 out of 465 pairs show statistically significant differences), **(b)** intermediate (247 out of 465 pairs) and big **(c)** (40 out of 465 pairs). The pairwise correlations are thus driven by large cells, and the cell volume can act as a confounding factor.

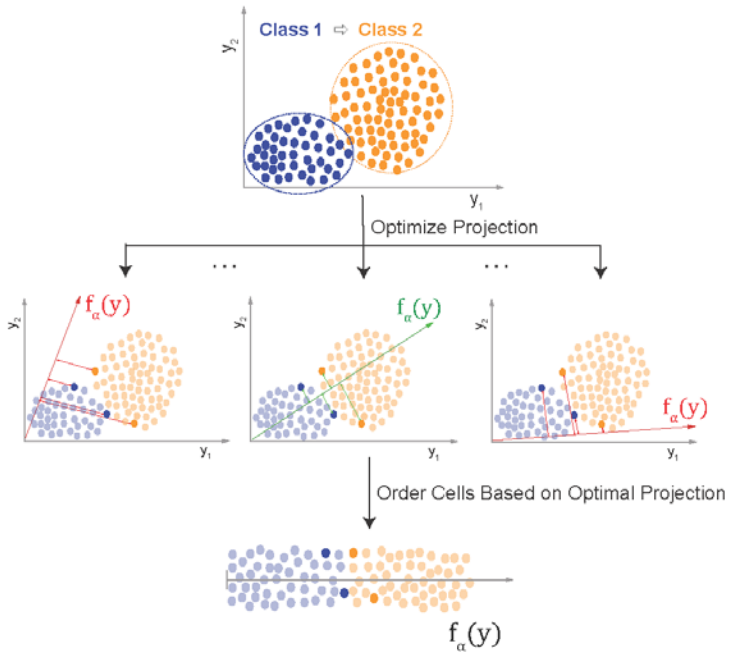



Supplementary Figure 5: **Cell cycle classification using decision trees.** **(a)** Measurements of the four cell cycle markers IdU, cyclin B1, p-HH3 and p-RB can be used for manual cell cycle phase assignment as described in [1]. Shown here are the results of this gating process for a THP-1 cell line, as performed in Cytobank. **(b)** Resulting structure of the decision tree after the training phase, with class labels and percentages indicated at the terminal nodes. The decision tree faithfully recapitulates the class assignment performed by manually gating, both in terms of the markers selected per class and the order of selection. **(c)** Cell cycle phase ratios in the three studied cell lines, as predicted by a decision tree trained on measurements of IdU, cyclin B1, p-HH3 and p-RB.



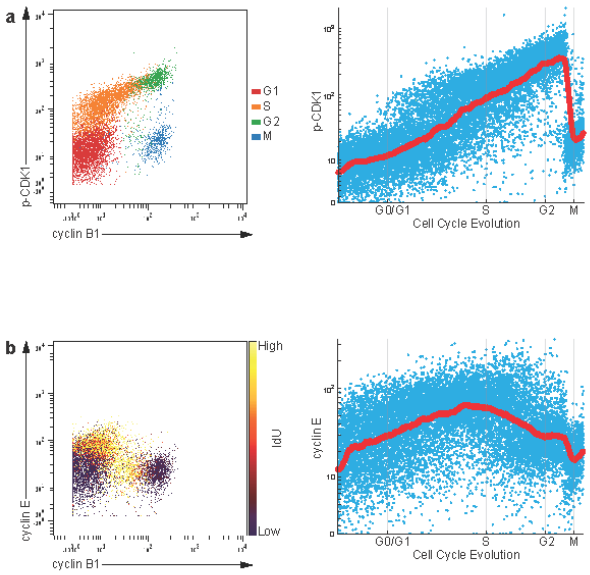

Supplementary Figure 6: **Decision trees and Gaussian mixture models enable automatic class assignment.** **(a)** Results of supervised learning with decision trees, in the space of hypothetical markers $y_1, y_2$ for two hypothetical classes 1 and 2. The decision boundaries are shown as areas with different colors. Class means (red circles), data covariance matrices and class proportions are used as initial parameters in a Gaussian mixture model with two components. **(b)** After fitting using the Expectation-Maximization (EM) [7] algorithm, the posterior probability that a single cell belongs to each class is returned and the class with the higher posterior is selected. Here, color indicates class assignment and opacity indicates posterior probability value. Refined class means are indicated as red circles.


Supplementary Figure 7: **The Gaussian Mixture Model improves cell cycle classification results of decision trees.** **(a)** Confusion matrix of the classification performance on an independent test set from HEK293T data after prediction using decision trees. **(b)** Final confusion matrix after classification refinement by fitting a Gaussian mixture model of four components to the same HEK293T data. The results indicate an overall improvement of the prediction across all classes when decision trees and GMMs are combined.



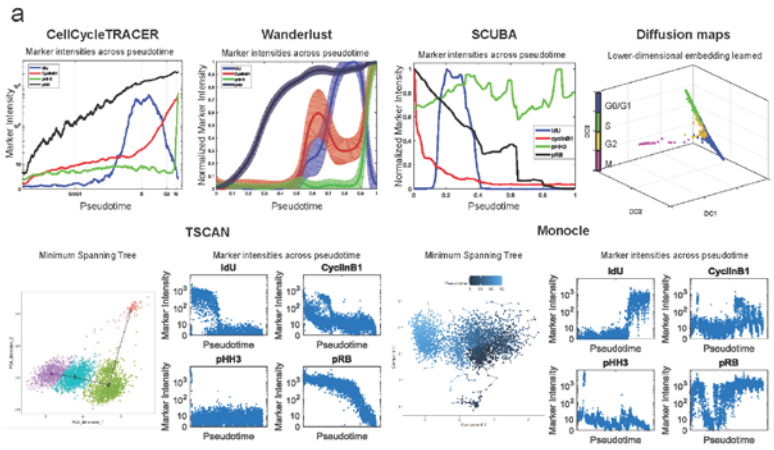

Supplementary Figure 8: **Outline of the trajectory reconstruction method.** Toy example of the trajectory reconstruction method for two classes in a two-dimensional space of hypothetical markers $y_1, y_2$. Given prior information about the single cell class labels and the class ordering, the best embedding $f_\alpha(y)$ is computed by selecting the function that optimally preserves the class ordering in the new subspace spanned by $f_\alpha(y)$. As class ordering violations are expected in the embedding due to noise or measurement outliers, we introduce slack variables, i.e. positive variables that penalize each constraint violation, and we minimize over the sum of all slack variables.

10




Supplementary Figure 9: **Cell cycle trajectories of p-CDK1 and cyclin E as independent validation.** **(a)** Phosphorylation of CDK1 on Tyr15 progressively increases during S and G2, peaks at the G2-M transition, and CDK1 is then dephosphorylated once cells enter M phase. Internal progressions of S phase and G2 phase are captured, and the G2-M transition is sharp. **(b)** At the same time, cyclin E progressively increases during G1, peaks at G1-S transition and degrades during S phase. In conclusion, the inferred cell cycle trajectories validate that our approach accurately captures S and G2 internal progression and the G2-M transition (p-CDK1), as well as the G1 phase progression and G1-S transition (cyclin E).

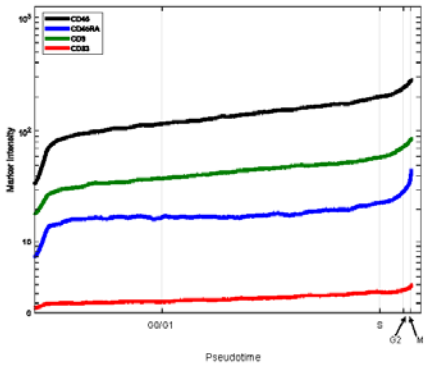
a



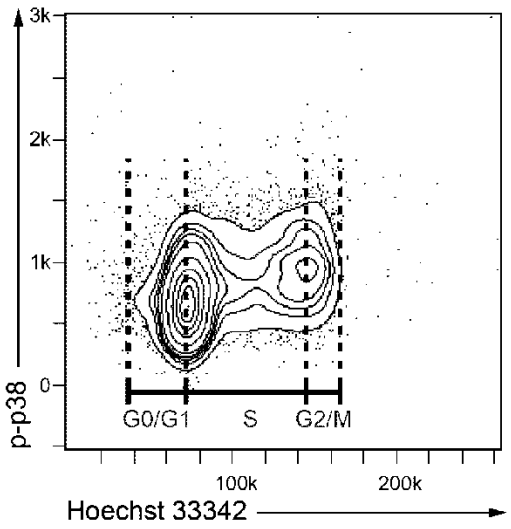

b

| | Implementations | Runtime (in seconds) |
|---|---|---|
| CellCycleTRACER | MATLAB | 9 |
| Wanderlust | MATLAB (CYT implementation) | 7 |
| SCUBA | MATLAB (with drtoolbox and princurve packages) | 627 |
| Diffusion Maps | MATLAB (estimated $\sigma$ = 0.16) | 101 |
| TSCAN | Bioconductor package | 120 |
| Monocle | Bioconductor package | 1344 |

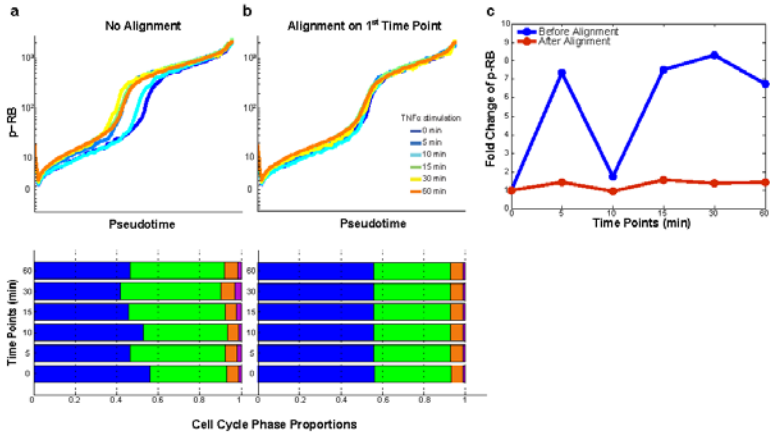
Supplementary Figure 10: **(a) Comparison of CellCycleTRACER with widely used trajectory reconstruction and embedding methods, as inferred from our four cell cycle marker measurements on a population of $n$ =3753 THP-1 cells.** The cell cycle fluctuations of the markers do not represent the underlying biology. Wanderlust [2] ordered the cells in the wrong order (G1→G2→S→M). SCUBA [6] reconstructed a G2→S→G0/G1 trajectory and incorporated M phase cells in the other clusters. TSCAN [4] reconstructed a M→S/G2→G0/G1 trajectory by mixing together G2 and S cells, and Monocle [8] ordered the data as G0/G1→M→G0/G1→G2→S, by ordering M phase cells in the middle of the G0/G1 cluster. Last, diffusion maps [3] constructed a non-linear, low-dimensional embedding of the data, which did not capture the known ordering. **(b)** Implementation details and runtimes of the above-mentioned methods.



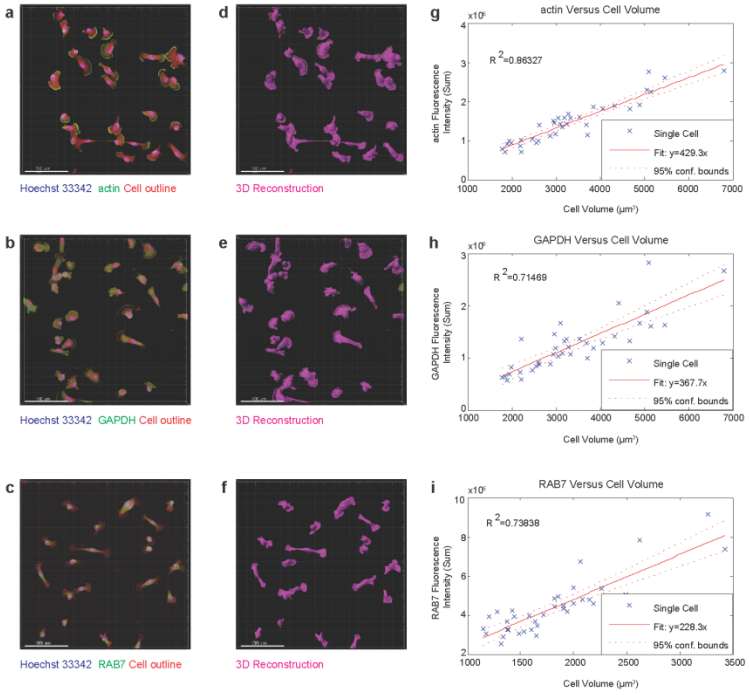

Supplementary Figure 11: **Cell cycle trajectories of cell surface markers.** Analysis of the published human T cell data [1] with CellCycleTRACER allowed the reconstruction of cell cycle trajectories of the cell surface markers. In agreement with the marker distributions across the cell cycle, shown in Supplementary Fig.2, the resulting trajectories for some of the surface proteins exhibited an increasing trend across the cell cycle, peaking as the cells enter M phase.



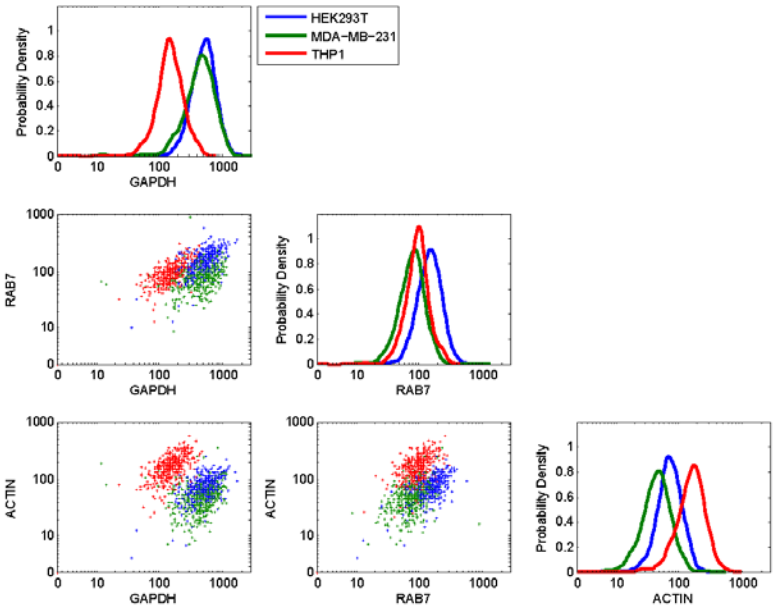

Supplementary Figure 12: **Cell-cycle-dependency of p38 phosphorylation in response to TNFα stimulation confirmed by flow cytometry.** THP-1 cells treated with TNFα for 15 min were measured by flow cytometry. A 1.5 fold increase of p-p38 level from the G0/G1 phase to G2/M phase can be observed, validating that the phosphorylation of p38 (Thr180/Tyr182) in response to TNFα stimulation is cell-cycle-dependent.



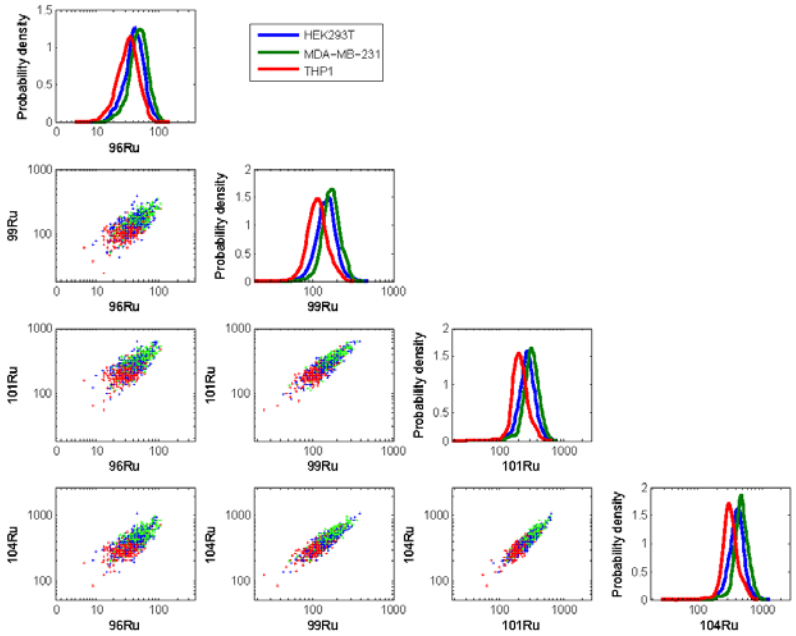

Supplementary Figure 13: **Trajectory alignment through cell cycle phase equalization removes duration-specific variability.** Trajectories of p-RB during TNF$\alpha$ stimulation in a population of HEK-293T cells, before and after cell cycle alignment. A larger population of S phase cells in time points 2, 4, 5, and 6 (**a**) results in relatively elevated levels of p-RB. When the cell cycle phases are aligned (**b**) no duration-specific variation is observed. Subsequently, estimating the fold change of the abundance of p-RB with respect to the control time point (**c**) results in large fold change values of time points 2, 4, 5, and 6; however, after equalizing the duration of the phases, p-RB levels are comparable across time points and the fold change is negligible.



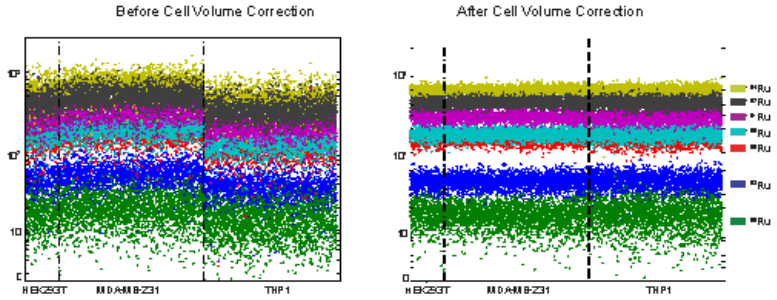

Supplementary Figure 14: **Validation of housekeeping proteins as a cell volume indicator. (a-c)** MDA-MB-231 cells were stained with Hoechst 33343 for the nucleus and Alexa Fluor 647 carboxylic acid succinimidyl ester for the cell outline. Housekeeping proteins actin, GAPDH and RAB7 were stained in (a), (b) and (c), respectively. **(d-f)** 3D reconstruction of corresponding images (a-c), using cell outline determined by Alexa Fluor 647 carboxylic acid succinimidyl ester. **(g-i)** Linear regression on the computed cell volume versus summed actin, GAPDH and RAB7 fluorescence intensity, respectively. Scale bar, 100$\mu$m.



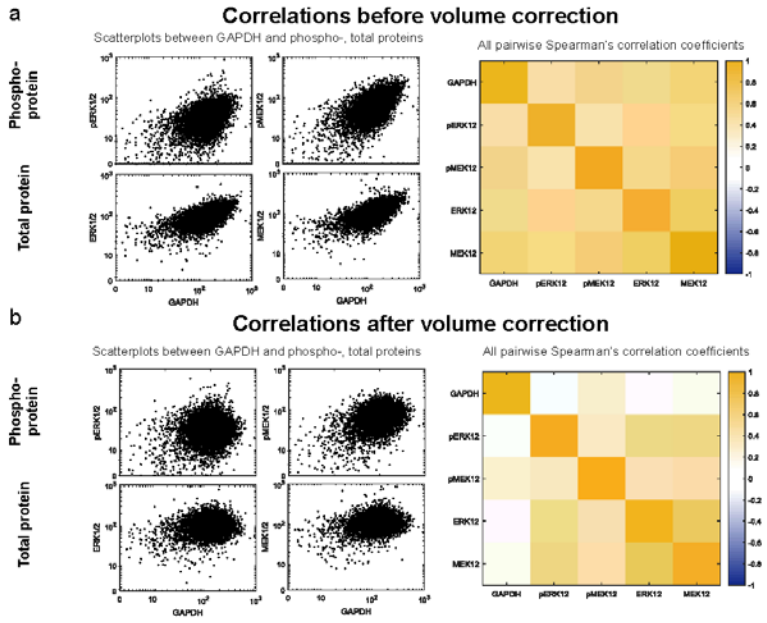

Supplementary Figure 15: **Housekeeping protein measurements across different cell lines reveal significant differences in protein abundances.** Probability density plots and pairwise scatter plots of the single cell measurements of housekeeping proteins GAPDH, RAB7 and actin in HEK293T, MDA-MB-231 and THP-1 cells show cell-line-specific variations in amounts of these proteins. For example, THP-1 cells have the lowest GAPDH levels but the highest actin levels. Thus, contrary to our expectations, housekeeping protein levels differ across cell types and are not optimal markers for cell volume correction.



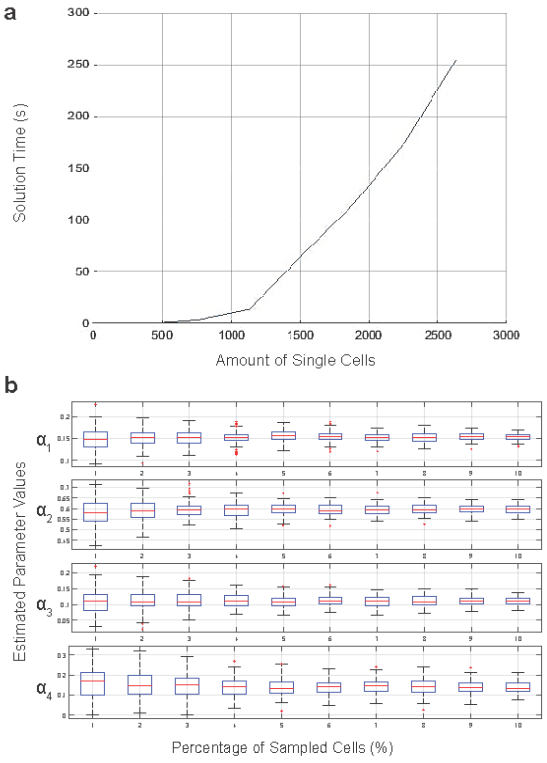

Supplementary Figure 16: **ASCQ Ru measurements are not affected by cell line variability.** Probability densities and pairwise scatter plots of the single cell measurements for four ruthenium stable isotopes, spanning all atomic masses, in HEK293T, MDA-MB-231 and THP-1 cells. In contrast to the housekeeping protein measurements, ASCQ_Ru is significantly more robust against cell-line-specific variations, allowing thus a more accurate quantification of cell volume.




Supplementary Figure 17: **Cell volume correction applied to ASCQ_Ru measurements removes cell type variability.** Data from all seven stable isotopes of ASCQ_Ru, indicated with different colors, as measured in in HEK293T, MDA-MB-231 and THP-1 cells (left). After cell volume correction (right), the measurements are aligned vertically and no cell-line-specific variation is observed.




Supplementary Figure 18: **Pairwise correlation coefficients between GAPDH and MEK1/2, ERK1/2 (total and activated state), before and after cell volume correction.** We assessed the abundance of both the total amount and the phosphorylated amount for proteins ERK1/2 and MEK1/2 and we subsequently examined whether and to which extent they are affected by volume. We used the ruthenium isotopes to normalize the data and correct for volume, and quantified the housekeeping protein GAPDH as an independent validation marker (GAPDH measurements were not corrected). **(a)** Before volume correction both ERK1/2 and MEK1/2 were correlated with GAPDH (Spearmans correlation coefficients shown on the left), in both their total and the activated state. **(b)** However, after cell volume correction the effect disappeared and correlation coefficients with GAPDH were significantly lower for all proteins in both states (total and active). Within-protein correlations were preserved, however.




Supplementary Figure 19: **Solution time and robustness of the linear programming (LP) optimization process.** (**a**) Runtime of the LP solution with respect to the number of single cells considered. (**b**) Boxplots of distributions of estimated values of parameters $\alpha$, acquired after 100 random repetitions of the optimization process, for a varying percentage of sampled single cells, indicates robust results when a minimum of 5% of the total population is sampled.

# Discussion

### Signaling network analysis at single-cell resolution

Signaling networks are centrally involved in the process of cellular information processing that controls cell functions and cell fate. Deregulated signaling network often leads to the emergence of disease. Recent advances in systems biology researchers have identified multiple layers of varieties, including genetic and non-genetic alterations, intrinsic and extinct factors that contribute to heterogeneous signaling network states and dynamics. Importantly, the essential role of signaling network heterogeneity in the initiation and development of diseases, such as cancer, has been revealed. Many recently developed techniques are now capable to quantify signaling events and network behaviors at the single-cell level.

Currently, more than 40 phosphorylation sites can be simultaneously quantified in mass cytometry-based single cell proteomics analysis (Bodenmiller et al., 2012; Lun et al., 2017). Imaging mass cytometry and several sequential imaging approaches now add the spatial information on top of the cellular signaling profiling (Giesen et al., 2014; Lin et al., 2015; Wang et al., 2017). These methods are ready to be used in systematical inference of signaling network behaviors and can be further applied in mechanistically modeling that explains varied signaling dynamics at the single-cell resolution (Hasenauer et al., 2014; Loos et al., 2017). Meanwhile, transcriptomics can be measured at the single-cell level to indicate activities of particular signaling pathways. Integrated with spatial information, transcriptomic methods are powerful in understanding paracrine signaling regulation in which secreted signaling proteins are largely involved (Lee et al., 2014; Shalek et al., 2014; Tang et al., 2009).

Technically, efforts have been made to increase the multiplexity of antibody-based single cell measurement: current advances in sequential imaging allows unlimited multiplexing (Wang et al., 2017). However, a caveat for these approaches is the antibody unavailability, especially for membrane-expressed receptor proteins and many intracellular phosphorylation sites. Single cell transcriptomic approaches are ready to assess the mRNA expression globally and in an unbiased manner. Although these methods are prone to biological and technical noise that drop-outs may exist in majority of cells for mRNAs that are not highly expressed, several approaches aim to computationally impute these drop-outs are available (Dijk et al., 2017; Li and Li, 2017; Ronen and Akalin, 2018).

Future questions for single-cell signaling characterization include the integration of multi-omics profiling to understand the signaling circuits as well as feedback controlling mechanisms between, for instance, signaling pathway activities and transcriptional programs. Applying available techniques using DNA-oligo-tagged antibodies (Peterson et al., 2017; Stoeckius et al., 2017), phosphorylation sites can be measured in combination with the whole transcriptomic sequencing in the same measured cells. Imaging mass cytometry now also allows simultaneous measurement of protein and RNA (Schulz et al., 2017) that can be further expanded to answer the questions regarding signaling controlling mechanisms between the phosphoprotein network and transcriptional regulation and the involvement of spatial factors, such as cell-to-cell contacts and protein localization, in such networks.

**Signaling protein abundance-dependent effects on signaling networks**

Here we describe the establishment and the thorough validation of an approach that measures influence of signaling protein overexpression on the signaling network states and dynamics in an abundance-dependent manner. Our approach applies mass cytometry-based single-cell analysis on samples with artificially induced protein expression variances that are achieved by transient transfection. Typically, in bulk measurement, transient transfection is considered problematic since it creates heterogeneous protein overexpression levels. Mass cytometry can resolve in the expression gradient and comprehensively profile the signaling network modulations specific to any range of expression titration. Thus, overexpression-induced network alternations can be studied as continuums.

Our newly developed approach has multiple layers of novelties. First, unlike most of previous signaling network studies which were done with cell lysates or in vitro kinase assays (Bodenmiller et al., 2010; Mok et al., 2010; Ochoa et al., 2016), we measured the kinase and phosphatase influence on signaling networks in undisrupted cells, addressing questions in a physiological context where the signaling network parameters can be interpreted at systems level. Second, for the first time, we studied overexpression-induced signaling network modulation; with the power of mass cytometry to preform single cell analysis, we quantified the protein abundance-dependency on network behaviors. Third, applying on a human kinome- and phosphatome-wide analysis, we showed our approach as screening approach that helped understanding, in global, the signaling effects induced by protein overexpression, and in screening for target proteins with specific influence on network states.

Using our method to study the key signaling proteins in EGFR network (Lun et al., 2017) and subsequently in a human kinome- and phosphatome-wide screen, we also revealed many novel biological insights. First, we discovered that overexpressed kinases had self-phosphorylation that, however, was often unable to induce downstream signaling activation, indicating proportion of phosphorylated proteins from the total to be deterministic for the downstream pathway activity. Second, for the first time, we revealed protein overexpression altered signaling dynamics that were measured as signaling amplitudes and peak times. This dynamic modulation is crucial for understanding oncogenic signaling caused by protein overexpression: the same concentration of extracellular stimulation might result in different signaling outcome if overexpression of certain kinases prolongs the pathway activation duration. Third, many kinases work in protein complex that they coordinate the process signaling transduction events. Our analysis showed that overexpression of a single subunit in a protein complex disrupted the complex function, likely due to competitive inhibition that typically resulted in dampened downstream signals. Fourth, we revealed that 54 kinases, when overexpressed, could induce EGF-independent ERK activation. Overexpression of these kinases was validated, in melanoma A375 cells, with the capability to induce resistance to a $BRAF^{V600E}$ inhibitor vemurafenib, suggesting that ligang-independent ERK activation to be a biomarker for drug resistance. Finally, we discovered that overexpression of ERK-specific phosphatases impacted the ERK-to-MEK negative feedback and prolonged MEK activation after EGF stimulation. In the context of ERK overexpression, prolonged MEK activity increased the duration of proliferative signal that could be an oncogenic-like signaling mechanism related to phosphatase overexpression which was previously unknown.

It also has to be noted that our analysis does not measure direct protein-protein interaction. Rather, we study the protein overexpression-induced network changes that can be resulted from direct or indirect effects (i.e., cell cycle- or stress-related signaling behaviors) that represent physiological cell signaling states under certain overexpression-induced perturbation. Moreover, unlike phosphoproteomics that globally analyzes phosphorylation sites on a signaling protein or in a signaling pathway, our antibody-based approach targets those pre-selected phosphorylation sites that are known to be critical in the cancer signaling networks. We do not measure or infer the absolute copy number of any measured protein or protein modification. Using antibodies to target proteins or protein modifications of interest, we compare the relative expression levels or phosphorylation variances between different samples.

**Perspective of abundance-dependency analysis**

By coupling mass cytometry-based single-cell analysis and transient transfection, our approach is powerful in the characterization of protein expression-dependency on signaling network responses. In general, this approach can be applied to answer questions in other fields of biological and biomedical researches that require protein abundance determination or quantification indicative for characterizing certain cellular phenotypes: for example, to quantify the molecular landscape at the decision point of stem cell differentiation or a phenotypical switch, such as epithelial to mesenchymal transition. In combination with the cell volume analysis method that is also developed in this thesis, the relative concentration information is ready to be determined for each analyzed protein. This reduces the single cell variances form the volume differences and is essential for understanding and modeling the cellular biological processes in which protein concentrations are deterministic for a biological outcome. Further potential technical developments include absolute quantification of overexpressed proteins at single-cell resolution by combining mass cytometry measurement with quantitative proteomics. Single cell protein absolute copy number information is essential for the development of realistic mechanistic modeling (Erickson et al., 2018).

In the analysis of kinase and phosphatase co-regulatory mechanism on signaling network behaviors, we have shown that the abundance-dependency analysis can be performed in a two-dimensional space that a combination of two proteins, one tagged by GFP and the other tagged by FLAG, are simultaneously overexpressed in cells, each with independent expression gradient. Further development following the same line may increase the mutliplexity of overexpressed proteins by applying other protein tags, such as GST, HA, or V5, or implementing PLAYR-based detection methods (Frei et al., 2016) for RNA barcode conjugated with the mRNA encoding analyzed protein. For biological processes requiring coordinative functions carried out by more than one protein such as the induction of iPS (induced pluripotent stem) cells (Takahashi and Yamanaka, 2006), this approach can characterize the multi-dimensional expression landscape for the cellular decision making. In cancer biology, a multiplexed overexpression assay may help identifying and quantifying critical intrinsic factors and their combination for acquired drug resistance.

The current version of our kinome and phosphatome overexpression library is constructed using a vector that encodes POIs under a CMV promotor. CMV promotor is typically used to achieve a relatively high expression level. In case the cell phenotypes of interest can be induced by low

expression level of a protein, the precise determination of protein abundance at the decision point can be performed by using a weak promotor, such as Ubiquitin C promoter (UBC) (Qin et al., 2010). Moreover, for proteins with high endogenous expression levels, CRISPR-based gene editing techniques (Mali et al., 2013) can be used to knock out gene endogenous expression before transient transfection.

In summary, the development of highly-multiplexed single-cell analysis techniques has allowed studying cell intracellular signaling networks at single-cell resolution. Here, we describe the development of a mass cytometry-based method that enables comprehensive characterization of protein abundance-dependent signaling network modulation and illustrate the use of our technique in a kinome- and phosphatome-wide analysis. This method can be potentially developed to determine the concentration of POIs at the cell fate decision point during development and in cancer cell transformation. Further technical development may increase the overexpression multiplexity for combinatorial screening. Elements in our expression vectors are also exchangeable to fit specific applications.

## Reference


Abbas, T., Keaton, M.A., and Dutta, A. (2013). Genomic instability in cancer. Cold Spring Harb. Perspect. Biol. *5*, a012914.

Abu-Remaileh, M., Bender, S., Raddatz, G., Ansari, I., Cohen, D., Gutekunst, J., Musch, T., Linhart, H., Breiling, A., Pikarsky, E., et al. (2015). Chronic inflammation induces a novel epigenetic program that is conserved in intestinal adenomas and in colorectal cancer. Cancer Res. *75*, 2120–2130.

Amir, E.D., Davis, K.L., Tadmor, M.D., Simonds, E.F., Levine, J.H., Bendall, S.C., Shenfeld, D.K., Krishnaswamy, S., Nolan, G.P., and Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat. Biotechnol. *31*, 545–552.

Angelo, M., Bendall, S.C., Finck, R., Hale, M.B., Hitzman, C., Borowsky, A.D., Levenson, R.M., Lowe, J.B., Liu, S.D., Zhao, S., et al. (2014). Multiplexed ion beam imaging of human breast tumors. Nat. Med. *20*, 436–442.

Aoki, K., Kumagai, Y., Sakurai, A., Komatsu, N., Fujita, Y., Shionyu, C., and Matsuda, M. (2013). Stochastic ERK activation induced by noise and cell-to-cell propagation regulates cell density-dependent proliferation. Mol. Cell *52*, 529–540.

Bandura, D.R., Baranov, V.I., Ornatsky, O.I., Antonov, A., Kinach, R., Lou, X., Pavlov, S., Vorobiev, S., Dick, J.E., and Tanner, S.D. (2009). Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal. Chem. *81*, 6813–6822.

Banerjee, A., and Gerondakis, S. (2007). Coordinating TLR-activated signaling pathways in cells of the immune system. Immunol. Cell Biol. *85*, 420–424.

Barrios-Rodiles, M., Brown, K.R., Ozdamar, B., Bose, R., Liu, Z., Donovan, R.S., Shinjo, F., Liu, Y., Dembowy, J., Taylor, I.W., et al. (2005). High-Throughput Mapping of a Dynamic Signaling Network in Mammalian Cells. Science (80-. ). *307*, 1621–1625.

Behbehani, G.K. (2018). Cell Cycle Analysis by Mass Cytometry. (Humana Press, New York, NY), pp. 105–124.

Behbehani, G.K., Bendall, S.C., Clutter, M.R., Fantl, W.J., and Nolan, G.P. (2012). Single-cell mass cytometry adapted to measurements of the cell cycle. Cytometry. A *81*, 552–566.

Bendall, S.C., Simonds, E.F., Qiu, P., Amir, E.D., Krutzik, P.O., Finck, R., Bruggner, R. V, Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science *332*, 687–696.

Bendall, S.C., Nolan, G.P., Roederer, M., and Chattopadhyay, P.K. (2012). A deep profiler's guide to cytometry. Trends Immunol. *33*, 323–332.

Bendall, S.C., Davis, K.L., Amir, E.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014). Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development. Cell *157*, 714–725.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B *57*, 289–300.

Bertoli, C., Skotheim, J.M., and de Bruin, R.A.M. (2013). Control of cell cycle transcription during G1 and S phases. Nat. Rev. Mol. Cell Biol. *14*, 518–528.

Bezbradica, J.S., and Medzhitov, R. (2009). Integration of cytokine and heterologous receptor signaling pathways. Nat. Immunol. *10*, 333–339.

Blanchette, F., Rivard, N., Rudd, P., Grondin, F., Attisano, L., and Dubois, C.M. (2001). Cross-talk between the p42/p44 MAP kinase and Smad pathways in transforming growth factor beta 1-induced furin gene transactivation. J. Biol. Chem. *276*, 33986–33994.

Bodenmiller, B. (2016). Multiplexed Epitope-Based Tissue Imaging for Discovery and Healthcare Applications. Cell Syst. *2*, 225–238.

Bodenmiller, B., Wanka, S., Kraft, C., Urban, J., Campbell, D., Pedrioli, P.G., Gerrits, B., Picotti, P., Lam, H., Vitek, O., et al. (2010). Phosphoproteomic Analysis Reveals Interconnected System-Wide Responses to Perturbations of Kinases and Phosphatases in Yeast. Sci. Signal. *3*, rs4-rs4.

Bodenmiller, B., Zunder, E.R., Finck, R., Chen, T.J., Savig, E.S., Bruggner, R. V, Simonds, E.F., Bendall, S.C., Sachs, K., Krutzik, P.O., et al. (2012). Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. Nat. Biotechnol. *30*, 858–867.

Bollu, L.R., Mazumdar, A., Savage, M.I., and Brown, P.H. (2017). Molecular Pathways: Targeting Protein Tyrosine Phosphatases in Cancer. Clin. Cancer Res. *23*, 2136–2142.

Bowman, T., Garcia, R., Turkson, J., and Jove, R. (2000). STATs in oncogenesis. Oncogene *19*, 2474–2488.

Bray, S.J. (2016). Notch signalling in context. Nat. Rev. Mol. Cell Biol. *17*, 722–735.

Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.-Y., Breitkreutz, B.-J., Stark, C., Liu, G., et al. (2010). A global protein kinase and phosphatase interaction network in yeast. Science *328*, 1043–1046.

Brownlie, R.J., and Zamoyska, R. (2013). T cell receptor signalling networks: branched, diversified and bounded. Nat. Rev. Immunol. *13*, 257–269.

Bruggner, R. V., Bodenmiller, B., Dill, D.L., Tibshirani, R.J., and Nolan, G.P. (2014). Automated identification of stratifying signatures in cellular subpopulations. Proc. Natl. Acad. Sci. *111*, E2770–E2777.

Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018). Mass-spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. bioRxiv 102681.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490.

Buenrostro, J.D., Corces, R., Wu, B., Schep, A.N., Lareau, C., Majeti, R., Chang, H., and Greenleaf, W. (2017). Single-cell epigenomics maps the continuous regulatory landscape of human hematopoietic differentiation. bioRxiv 109843.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat. Biotechnol. *33*, 155–160.

Bunt, G., and Wouters, F.S. (2017). FRET from single to multiplexed signaling events. Biophys. Rev. *9*, 119–129.

Burack, W.R., and Shaw, A.S. (2005). Live Cell Imaging of ERK and MEK: simple binding equilibrium explains the regulated nucleocytoplasmic distribution of ERK. J. Biol. Chem. *280*, 3832–3837.

Burgermeister, E., and Seger, R. (2008). PPAR γ and MEK Interactions in Cancer. PPAR Res. *2008*, 1–16.

Bywater, M.J., Pearson, R.B., McArthur, G.A., and Hannan, R.D. (2013). Dysregulation of the basal RNA polymerase transcription apparatus in cancer. Nat. Rev. Cancer *13*, 299–314.

Cagnol, S., and Chambard, J.-C. (2010). ERK and cell death: Mechanisms of ERK-induced cell death - apoptosis, autophagy and senescence. FEBS J. *277*, 2–21.

Cardaci, S., Filomeni, G., and Ciriolo, M.R. (2012). Redox implications of AMPK-mediated signal transduction beyond energetic clues. J. Cell Sci. *125*, 2115–2125.

Castedo, M., Perfettini, J.-L., Roumier, T., and Kroemer, G. (2002). Cyclin-dependent kinase-1: linking apoptosis to cell cycle and mitotic catastrophe. Cell Death Differ. *9*, 1287–1293.

Cepero, V., Sierra, J.R., Corso, S., Ghiso, E., Casorzo, L., Perera, T., Comoglio, P.M., and Giordano, S. (2010). MET and KRAS gene amplification mediates acquired resistance to MET tyrosine kinase inhibitors. Cancer Res. *70*, 7580–7590.

Chan, T.E., Stumpf, M.P.H., and Babtie, A.C. (2017). Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. Cell Syst. *5*, 251–267.e3.

Chaube, B., Malvi, P., Singh, S. V, Mohammad, N., Viollet, B., and Bhat, M.K. (2015). AMPK maintains energy homeostasis and survival in cancer cells via regulating p38/PGC-1α-mediated mitochondrial biogenesis. Cell Death Discov. *1*, 15063.

Chaudhuri, O., Koshy, S.T., Branco da Cunha, C., Shin, J.-W., Verbeke, C.S., Allison, K.H., and Mooney, D.J. (2014). Extracellular matrix stiffness and composition jointly regulate the induction of malignant phenotypes in mammary epithelium. Nat. Mater. *13*, 970–978.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., Zhuang, X., Crosetto, N., Bienko, M., Oudenaarden, A. van, Femino, A.M., Fay, F.S., et al. (2015). RNA imaging. Spatially resolved,

highly multiplexed RNA profiling in single cells. Science *348*, aaa6090.

Chen, M.J., Dixon, J.E., and Manning, G. (2017). Genomics and evolution of protein phosphatases. Sci. Signal. *10*, eaag1796.

Chevrier, S., Crowell, H., Zanotelli, V.R.T., Engler, S., Robinson, M.D., and Bodenmiller, B. (2017a). Channel crosstalk correction in suspension and imaging mass cytometry. bioRxiv 185744.

Chevrier, S., Levine, J.H., Zanotelli, V.R.T., Silina, K., Schulz, D., Bacac, M., Ries, C.H., Ailles, L., Jewett, M.A.S., Moch, H., et al. (2017b). An Immune Atlas of Clear Cell Renal Cell Carcinoma. Cell *169*, 736–749.e18.

Citri, A., and Yarden, Y. (2006). EGF-ERBB signalling: towards the systems level. Nat. Rev. Mol. Cell Biol. *7*, 505–516.

Clapham, D.E. (2007). Calcium signaling. Cell *131*, 1047–1058.

Clevers, H., and Nusse, R. (2012). Wnt/β-catenin signaling and disease. Cell *149*, 1192–1205.

Cohen, P., and Frame, S. (2001). The renaissance of GSK3. Nat. Rev. Mol. Cell Biol. *2*, 769–776.

Corbalán-García, S., and Gómez-Fernández, J.C. (2006). Protein kinase C regulatory domains: The art of decoding many different signals in membranes. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids *1761*, 633–654.

Corcoran, R.B., Cheng, K.A., Hata, A.N., Faber, A.C., Ebi, H., Coffee, E.M., Greninger, P., Brown, R.D., Godfrey, J.T., Cohoon, T.J., et al. (2013). Synthetic lethal interaction of combined BCL-XL and MEK inhibition promotes tumor regressions in KRAS mutant cancer models. Cancer Cell *23*, 121–128.

Couzens, A.L., Knight, J.D.R., Kean, M.J., Teo, G., Weiss, A., Dunham, W.H., Lin, Z.-Y., Bagshaw, R.D., Sicheri, F., Pawson, T., et al. (2013). Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. Sci. Signal. *6*, rs15.

Creixell, P., Schoof, E.M., Simpson, C.D., Longden, J., Miller, C.J., Lou, H.J., Perryman, L., Cox, T.R., Zivanovic, N., Palmeri, A., et al. (2015). Kinome-wide decoding of network-attacking

mutations rewiring cancer signaling. Cell *163*, 202–217.

Cuenda, A., and Rousseau, S. (2007). p38 MAP-Kinases pathway regulation, function and role in human diseases. Biochim. Biophys. Acta - Mol. Cell Res. *1773*, 1358–1375.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910–914.

Davies, H., Bignell, G.R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M.J., Bottomley, W., et al. (2002). Mutations of the BRAF gene in human cancer. Nature *417*, 949–954.

Davis, R.J. (2000). Signal Transduction by the JNK Group of MAP Kinases. Cell *103*, 239–252.

Derynck, R., Muthusamy, B.P., and Saeteurn, K.Y. (2014). Signaling pathway cooperation in TGF-β-induced epithelial–mesenchymal transition. Curr. Opin. Cell Biol. *31*, 56–66.

Dijk, D. van, Nainys, J., Sharma, R., Kathail, P., Carr, A.J., Moon, K.R., Mazutis, L., Wolf, G., Krishnaswamy, S., and Pe'er, D. (2017). MAGIC: A diffusion-based imputation method reveals gene-gene interactions in single-cell RNA-sequencing data. bioRxiv 111591.

Donati, G., and Watt, F.M. (2015). Stem Cell Heterogeneity and Plasticity in Epithelia. Cell Stem Cell *16*, 465–476.

Du, W., and Elemento, O. (2015). Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. Oncogene *34*, 3215–3225.

Duan, Y., Wong, W., Chua, S.C., Wee, H.L., Lim, S.G., Chua, B.T., and Ho, H.K. (2016). Overexpression of Tyro3 and its implications on hepatocellular carcinoma progression. Int. J. Oncol. *48*, 358–366.

Eralp, Y., Derin, D., Ozluk, Y., Yavuz, E., Guney, N., Saip, P., Muslumanoglu, M., Igci, A., Kücücük, S., Dincer, M., et al. (2008). MAPK overexpression is associated with anthracycline resistance and increased risk for recurrence in patients with triple-negative breast cancer. Ann. Oncol. *19*, 669–674.

Erickson, K.E., Rukhlenko, O.S., Posner, R.G., Hlavacek, W.S., and Kholodenko, B.N. (2018). New insights into RAS biology reinvigorate interest in mathematical modeling of RAS signaling.

Semin. Cancer Biol.

Feinberg, A.P. (2007). Phenotypic plasticity and the epigenetics of human disease. Nature *447*, 433–440.

Finck, R., Simonds, E.F., Jager, A., Krishnaswamy, S., Sachs, K., Fantl, W., Pe'er, D., Nolan, G.P., and Bendall, S.C. (2013). Normalization of mass cytometry data with bead standards. Cytometry. A *83*, 483–494.

Fisher, D., Krasinska, L., Coudreuse, D., Novák, B., Rempel, R.E., Adamczewski, J., Maller, J.L., Hunt, T., and Blow, J.J. (2012). Phosphorylation network dynamics in the control of cell cycle transitions. J. Cell Sci. *125*, 4703–4711.

Fleuren, E.D.G., Zhang, L., Wu, J., and Daly, R.J. (2016). The kinome "at large" in cancer. Nat. Rev. Cancer *16*, 83–98.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., et al. (2012). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. *41*, D808–D815.

Frechin, M., Stoeger, T., Daetwyler, S., Gehin, C., Battich, N., Damm, E.-M., Stergiou, L., Riezman, H., and Pelkmans, L. (2015). Cell-intrinsic adaptation of lipid composition to local crowding drives social behaviour. Nature *523*, 88–91.

Frei, A.P., Bava, F.A., Zunder, E.R., Hsieh, E.W.Y., Chen, S.Y., Nolan, G.P., and Gherardini, P.F. (2016). Highly multiplexed simultaneous detection of RNAs and proteins in single cells. Nat. Methods *13*.

Garsed, D.W., Marshall, O.J., Corbin, V.D.A., Hsu, A., Di Stefano, L., Schröder, J., Li, J., Feng, Z.-P., Kim, B.W., Kowarsky, M., et al. (2014). The architecture and evolution of cancer neochromosomes. Cancer Cell *26*, 653–667.

Geißler, D., Stufler, S., Löhmannsröben, H.-G., and Hildebrandt, N. (2013). Six-Color Time-Resolved Förster Resonance Energy Transfer for Ultrasensitive Multiplexed Biosensing. J. Am. Chem. Soc. *135*, 1102–1109.

Gerdes, M.J., Sevinsky, C.J., Sood, A., Adak, S., Bello, M.O., Bordwell, A., Can, A., Corwin, A., Dinn, S., Filkins, R.J., et al. (2013). Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. Proc. Natl. Acad. Sci. U. S. A. *110*, 11982–11987.

Giesen, C., Wang, H.A.O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P.J., Grolimund, D., Buhmann, J.M., Brandt, S., et al. (2014). Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat. Methods *11*, 417–422.

Goh, L.K., and Sorkin, A. (2013). Endocytosis of receptor tyrosine kinases. Cold Spring Harb. Perspect. Biol. *5*, a017459.

Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vasquez, G., and Nolan, G. (2017). Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. bioRxiv 203166.

Gonzalo, S. (2010). Epigenetic alterations in aging. J. Appl. Physiol. *109*, 586–597.

Govindarajan, B., Sligh, J.E., Vincent, B.J., Li, M., Canter, J.A., Nickoloff, B.J., Rodenburg, R.J., Smeitink, J.A., Oberley, L., Zhang, Y., et al. (2007). Overexpression of Akt converts radial growth melanoma to vertical growth melanoma. J. Clin. Invest. *117*, 719–729.

Graf, J.F., and Zavodszky, M.I. (2017). Characterizing the heterogeneity of tumor tissues from spatially resolved molecular measures. PLoS One *12*, e0188878.

Grecco, H.E., Schmick, M., and Bastiaens, P.I.H. (2011). Signaling from the Living Plasma Membrane. Cell *144*, 897–909.

Groves, J.T., and Kuriyan, J. (2010). Molecular mechanisms in signal transduction at the membrane. Nat. Struct. Mol. Biol. *17*, 659–665.

Guan, K.L., and Butch, E. (1995). Isolation and characterization of a novel dual specific phosphatase, HVH2, which selectively dephosphorylates the mitogen-activated protein kinase. J. Biol. Chem. *270*, 7197–7203.

Guo, J., Grow, E.J., Yi, C., Mlcochova, H., Maher, G.J., Lindskog, C., Murphy, P.J., Wike, C.L., Carrell, D.T., Goriely, A., et al. (2017). Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development. Cell Stem Cell *21*, 533–546.e6.

Gut, G., Tadmor, M.D., Pe'er, D., Pelkmans, L., and Liberali, P. (2015). Trajectories of cell-cycle progression from fixed cell populations. Nat. Methods *12*, 951–954.

Hagberg, A.A., Schult, D.A., and Swart, P.J. (2008). Exploring Network Structure, Dynamics,

and Function using NetworkX. Proc. 7th Python Sci. Conf.

Han, T., Xiang, D.-M., Sun, W., Liu, N., Sun, H.-L., Wen, W., Shen, W.-F., Wang, R.-Y., Chen, C., Wang, X., et al. (2015). PTPN11/Shp2 Overexpression Enhances Liver Cancer Progression and Predicts Poor Prognosis of Patients. J. Hepatol.

Hasenauer, J., Hasenauer, C., Hucho, T., and Theis, F.J. (2014). ODE Constrained Mixture Modelling: A Method for Unraveling Subpopulation Structures and Dynamics. PLoS Comput. Biol. *10*, e1003686.

Hayashida, T., Decaestecker, M., and Schnaper, H.W. (2003). Cross-talk between ERK MAP kinase and Smad signaling pathways enhances TGF-beta-dependent responses in human mesangial cells. FASEB J. *17*, 1576–1578.

Heinrich, R., Neel, B.G., and Rapoport, T.A. (2002). Mathematical Models of Protein Kinase Signal Transduction. Mol. Cell *9*, 957–970.

Hendriks, R.W., Yuvaraj, S., and Kil, L.P. (2014). Targeting Bruton's tyrosine kinase in B cell malignancies. Nat. Rev. Cancer *14*, 219–232.

Hennessy, B.T., Smith, D.L., Ram, P.T., Lu, Y., and Mills, G.B. (2005). Exploiting the PI3K/AKT Pathway for Cancer Drug Discovery. Nat. Rev. Drug Discov. *4*, 988–1004.

Hetz, C., and Saxena, S. (2017). ER stress and the unfolded protein response in neurodegeneration. Nat. Rev. Neurol. *13*, 477–491.

Holland, A.J., and Cleveland, D.W. (2012). Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements. Nat. Med. *18*, 1630–1638.

Hoppe, A.D., Scott, B.L., Welliver, T.P., Straight, S.W., and Swanson, J.A. (2013). N-Way FRET Microscopy of Multiple Protein-Protein Interactions in Live Cells. PLoS One *8*, e64760.

Horgan, R.P., and Kenny, L.C. (2011). "Omic" technologies: genomics, transcriptomics, proteomics and metabolomics. Obstet. Gynaecol. *13*, 189–195.

Horn, T., Sandmann, T., Fischer, B., Axelsson, E., Huber, W., and Boutros, M. (2011). Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. Nat. Methods *8*, 341–346.

Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015).

PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. *43*, D512–D520.

Huang, X.-T., Zhu, Y., Hang Chan, L.L., Zhao, Z., and Yan, H. (2016). Inference of cellular level signaling networks using single-cell gene expression data in *C. elegans* reveals mechanisms of cell fate specification. Bioinformatics *33*, btw796.

Hughes, A.J., Spelke, D.P., Xu, Z., Kang, C.-C., Schaffer, D. V, and Herr, A.E. (2014). Single-cell western blotting. Nat. Methods *11*, 749–755.

Hughey, J.J., Lee, T.K., and Covert, M.W. (2009). Computational modeling of mammalian signaling networks. Wiley Interdiscip. Rev. Syst. Biol. Med. *2*, 194–209.

Jeffrey, K.L., Camps, M., Rommel, C., and Mackay, C.R. (2007). Targeting dual-specificity phosphatases: manipulating MAP kinase signalling and immune responses. Nat. Rev. Drug Discov. *6*, 391–403.

Jeong, H.M., Kwon, M.J., and Shin, Y.K. (2014). Overexpression of Cancer-Associated Genes via Epigenetic Derepression Mechanisms in Gynecologic Cancer. Front. Oncol. *4*, 12.

Johannessen, C.M., Boehm, J.S., Kim, S.Y., Thomas, S.R., Wardwell, L., Johnson, L.A., Emery, C.M., Stransky, N., Cogdill, A.P., Barretina, J., et al. (2010). COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. Nature *468*, 968–972.

Julien, S.G., Dubé, N., Read, M., Penney, J., Paquet, M., Han, Y., Kennedy, B.P., Muller, W.J., and Tremblay, M.L. (2007). Protein tyrosine phosphatase 1B deficiency or inhibition delays ErbB2-induced mammary tumorigenesis and protects from lung metastasis. Nat. Genet. *39*, 338–346.

Julien, S.G., Dubé, N., Hardy, S., and Tremblay, M.L. (2011). Inside the human cancer tyrosine phosphatome. Nat. Rev. Cancer *11*, 35–49.

Jungmann, R., Avendaño, M.S., Woehrstein, J.B., Dai, M., Shih, W.M., and Yin, P. (2014). Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. Nat. Methods *11*, 313–318.

Kellogg, R.A., and Tay, S. (2015). Noise Facilitates Transcriptional Control under Dynamic Inputs. Cell *160*, 381–392.

Kim, D., Rath, O., Kolch, W., and Cho, K.-H. (2007). A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK pathways. Oncogene *26*, 4571–4579.

Kim, N.-G., Koh, E., Chen, X., and Gumbiner, B.M. (2011). E-cadherin mediates contact inhibition of proliferation through Hippo signaling-pathway components. Proc. Natl. Acad. Sci. U. S. A. *108*, 11930–11935.

Kim, S.Y., Jeong, S., Jung, E., Baik, K.-H., Chang, M.H., Kim, S.A., Shim, J.-H., Chun, E., and Lee, K.-Y. (2012). AMP-activated protein kinase-α1 as an activating kinase of TGF-β-activated kinase 1 has a key role in inflammatory signals. Cell Death Dis. *3*, e357.

Koseska, A., and Bastiaens, P.I. (2017). Cell signaling as a cognitive process. EMBO J. e201695383.

Kowalczyk, M.S., Tirosh, I., Heckl, D., Rao, T.N., Dixit, A., Haas, B.J., Schneider, R.K., Wagers, A.J., Ebert, B.L., and Regev, A. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res. *25*, 1860–1872.

Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. Phys. Rev. E *69*, 66138.

Krishnaswamy, S., Spitzer, M.H., Mingueneau, M., Bendall, S.C., Litvin, O., Stone, E., Pe'er, D., and Nolan, G.P. (2014). Conditional density-based analysis of T cell signaling in single-cell data. Science science.1250689-.

Krishnaswamy, S., Zivanovic, N., Sharma, R., Pe'er, D., and Bodenmiller, B. (2017). Learning Edge Rewiring in EMT from Single Cell Data. bioRxiv 155028.

Krumsiek, J., Suhre, K., Illig, T., Adamski, J., and Theis, F.J. (2011). Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. BMC Syst. Biol. *5*, 21.

Krutzik, P.O., and Nolan, G.P. (2006). Fluorescent cell barcoding in flow cytometry allows high-throughput drug screening and signaling profiling. Nat. Methods *3*, 361–368.

Krutzik, P.O., Crane, J.M., Clutter, M.R., and Nolan, G.P. (2008). High-content single-cell drug screening with phosphospecific flow cytometry. Nat. Chem. Biol. *4*, 132–142.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and

Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Kuchenov, D., Laketa, V., Stein, F., Salopiata, F., Klingmüller, U., and Schultz, C. (2016). High-Content Imaging Platform for Profiling Intracellular Signaling Network Activity in Living Cells. Cell Chem. Biol. *23*, 1550–1559.

Lane, K., Van Valen, D., DeFelice, M.M., Macklin, D.N., Kudo, T., Jaimovich, A., Carr, A., Meyer, T., Pe'er, D., Boutet, S.C., et al. (2017). Measuring Signaling and RNA-Seq in the Same Cell Links Gene Expression to Dynamic Patterns of NF-κB Activation. Cell Syst. *4*, 458–469.e5.

Lavoie, H., and Therrien, M. (2015). Regulation of RAF protein kinases in ERK signalling. Nat. Rev. Mol. Cell Biol. *16*, 281–298.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. Science *343*, 1360–1363.

Lee, T.K., Denny, E.M., Sanghvi, J.C., Gaston, J.E., Maynard, N.D., Hughey, J.J., and Covert, M.W. (2009). A noisy paracrine signal determines the cellular NF-kappaB response to lipopolysaccharide. Sci. Signal. *2*, ra65.

Lee, Y.-K., Hwang, J.-T., Kwon, D.Y., Surh, Y.-J., and Park, O.J. (2010). Induction of apoptosis by quercetin is mediated through AMPKalpha1/ASK1/p38 pathway. Cancer Lett. *292*, 228–236.

Lemmon, M.A., and Schlessinger, J. (2010). Cell Signaling by Receptor Tyrosine Kinases. Cell *141*, 1117–1134.

Li, W.V., and Li, J.J. (2017). scImpute: Accurate And Robust Imputation For Single Cell RNA-Seq Data. bioRxiv 141598.

Li, C., Ma, H., Wang, Y., Cao, Z., Graves-Deal, R., Powell, A.E., Starchenko, A., Ayers, G.D., Washington, M.K., Kamath, V., et al. (2014). Excess PLAC8 promotes an unconventional ERK2-dependent EMT in colon cancer. J. Clin. Invest. *124*, 2172–2187.

Lidke, D.S., Huang, F., Post, J.N., Rieger, B., Wilsbacher, J., Thomas, J.L., Pouysségur, J., Jovin, T.M., and Lenormand, P. (2010). ERK nuclear translocation is dimerization-independent but controlled by the rate of phosphorylation. J. Biol. Chem. *285*, 3092–3102.

Lien, W.-H., and Fuchs, E. (2014). Wnt some lose some: transcriptional governance of stem cells by Wnt/β-catenin signaling. Genes Dev. *28*, 1517–1532.

Lin, J.-R., Fallahi-Sichani, M., and Sorger, P.K. (2015). Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. Nat. Commun. *6*, 8390.

Linding, R., Jensen, L.J., Ostheimer, G.J., van Vugt, M.A.T.M., Jørgensen, C., Miron, I.M., Diella, F., Colwill, K., Taylor, L., Elder, K., et al. (2007). Systematic discovery of in vivo phosphorylation networks. Cell *129*, 1415–1426.

Little, A.S., Balmanno, K., Sale, M.J., Newman, S., Dry, J.R., Hampson, M., Edwards, P.A.W., Smith, P.D., and Cook, S.J. (2011). Amplification of the driving oncogene, KRAS or BRAF, underpins acquired resistance to MEK1/2 inhibitors in colorectal cancer cells. Sci. Signal. *4*, ra17.

Liu, K., Chen, H.-L., Wang, S., Gu, M.-M., Chen, X.-M., Zhang, S.-L., Yu, K.-J., and You, Q.-S. (2016). High Expression of RIOK2 and NOB1 Predict Human Non-small Cell Lung Cancer Outcomes. Sci. Rep. *6*, 28666.

Logue, J.S., and Morrison, D.K. (2012). Complexity in the signaling network: insights from the use of targeted inhibitors in cancer therapy. Genes Dev. *26*, 641–650.

Loos, C., Moeller, K., Fröhlich, F., Hucho, T., and Hasenauer, J. (2017). Mechanistic hierarchical population model identifies latent causes of cell-to-cell variability. bioRxiv 171561.

De Los Angeles, A., Ferrari, F., Xi, R., Fujiwara, Y., Benvenisty, N., Deng, H., Hochedlinger, K., Jaenisch, R., Lee, S., Leitch, H.G., et al. (2015). Hallmarks of pluripotency. Nature *525*, 469–478.

Low, H.B., and Zhang, Y. (2016). Regulatory Roles of MAPK Phosphatases in Cancer. Immune Netw. *16*, 85.

Lun, X.-K., Zanotelli, V.R.T., Wade, J.D., Schapiro, D., Tognetti, M., Dobberstein, N., and Bodenmiller, B. (2017). Influence of node abundance on signaling network state and dynamics analyzed by mass cytometry. Nat. Biotechnol.

Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Manning, B.D., and Cantley, L.C. (2007). AKT/PKB signaling: navigating downstream. Cell *129*, 1261–1274.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science *298*, 1912–1934.

Marguerat, S., and Bähler, J. (2012). Coordinating genome expression with cell size. Trends Genet. *28*, 560–565.

Marin, T.M., Clemente, C.F.M.Z., Santos, A.M., Picardi, P.K., Pascoal, V.D.B., Lopes-Cendes, I., Saad, M.J.A., and Franchini, K.G. (2008). Shp2 negatively regulates growth in cardiomyocytes by controlling focal adhesion kinase/Src and mTOR pathways. Circ. Res. *103*, 813–824.

Massague, J. (2003). Integration of Smad and MAPK pathways: a link and a linker revisited. Genes Dev. *17*, 2993–2997.

Massagué, J. (2012). TGFβ signalling in context. Nat. Rev. Mol. Cell Biol. *13*, 616–630.

Mattila, E., Pellinen, T., Nevo, J., Vuoriluoto, K., Arjonen, A., and Ivaska, J. (2005). Negative regulation of EGFR signalling through integrin-α1β1-mediated activation of protein tyrosine phosphatase TCPTP. Nat. Cell Biol. *7*, 78–85.

McDavid, A., Finak, G., and Gottardo, R. (2016). The contribution of cell cycle to heterogeneity in single-cell RNA-seq data. Nat. Biotechnol. *34*, 591–593.

McGranahan, N., and Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell *168*, 613–628.

Meek, D.W. (2009). Tumour suppression by p53: a role for the DNA damage response? Nat. Rev. Cancer *9*, 714–723.

Mendoza, M.C., Er, E.E., and Blenis, J. (2011). The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. Trends Biochem. Sci. *36*, 320–328.

Mihaylova, M.M., and Shaw, R.J. (2011). The AMPK signalling pathway coordinates cell growth, autophagy and metabolism. Nat. Cell Biol. *13*, 1016–1023.

Mitra, S.K., Hanson, D.A., and Schlaepfer, D.D. (2005). Focal adhesion kinase: in command and control of cell motility. Nat. Rev. Mol. Cell Biol. *6*, 56–68.

Mok, J., Kim, P.M., Lam, H.Y.K., Piccirillo, S., Zhou, X., Jeschke, G.R., Sheridan, D.L., Parker, S.A., Desai, V., Jwa, M., et al. (2010). Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. Sci. Signal. *3*, ra12.

Newton, K., and Dixit, V.M. (2012). Signaling in innate immunity and inflammation. Cold Spring Harb. Perspect. Biol. *4*, a006049.

Ng, A.H.C., Dean Chamberlain, M., Situ, H., Lee, V., and Wheeler, A.R. (2015). Digital microfluidic immunocytochemistry in single cells. Nat. Commun. *6*, 7513.

Nielsen, M.M.B., Lambertsen, K.L., Clausen, B.H., Meyer, M., Bhandari, D.R., Larsen, S.T., Poulsen, S.S., Spengler, B., Janfelt, C., and Hansen, H.S. (2016). Mass spectrometry imaging of biomarker lipids for phagocytosis and signalling during focal cerebral ischaemia. Sci. Rep. *6*, 39571.

Niewiadomski, P., Kong, J.H., Ahrends, R., Ma, Y., Humke, E.W., Khan, S., Teruel, M.N., Novitch, B.G., and Rohatgi, R. (2014). Gli Protein Activity Is Controlled by Multisite Phosphorylation in Vertebrate Hedgehog Signaling. Cell Rep. *6*, 168–181.

Nishi, H., Demir, E., and Panchenko, A.R. (2015). Crosstalk between Signaling Pathways Provided by Single and Multiple Protein Phosphorylation Sites. J. Mol. Biol. *427*, 511–520.

Nyati, M.K., Morgan, M.A., Feng, F.Y., and Lawrence, T.S. (2006). Integration of EGFR inhibitors with radiochemotherapy. Nat. Rev. Cancer *6*, 876–885.

Ochoa, D., Jonikas, M., Lawrence, R.T., El Debs, B., Selkrig, J., Typas, A., Villén, J., Santos, S.D., and Beltrao, P. (2016). An atlas of human kinase regulation. Mol. Syst. Biol. *12*, 888.

Olayioye, M.A., Neve, R.M., Lane, H.A., and Hynes, N.E. (2000). The ErbB signaling network: receptor heterodimerization in development and cancer. EMBO J. *19*, 3159–3167.

Oliva, J.L., Griner, E.M., and Kazanietz, M.G. (2005). PKC isozymes and diacylglycerol-regulated proteins as effectors of growth factor receptors. Growth Factors *23*, 245–252.

Padovan-Merhar, O., Nair, G.P., Biaesch, A.G., Mayer, A., Scarfone, S., Foley, S.W., Wu, A.R., Churchman, L.S., Singh, A., Raj, A., et al. (2015). Single Mammalian Cells Compensate for

Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms. Mol. Cell *58*, 339–352.

Di Palma, S., and Bodenmiller, B. (2014). Unraveling cell populations in tumors by single-cell mass cytometry. Curr. Opin. Biotechnol.

Parsons, S.J., and Parsons, J.T. (2004). Src family kinases, key regulators of signal transduction. Oncogene *23*, 7906–7909.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B. V, Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

Paul, M.K., and Mukhopadhyay, A.K. (2004). Tyrosine kinase - Role and significance in Cancer. Int. J. Med. Sci. *1*, 101–115.

Pawson, T., and Warner, N. (2007). Oncogenic re-wiring of cellular signaling pathways. Oncogene *26*, 1268–1275.

Pennock, N.D., White, J.T., Cross, E.W., Cheney, E.E., Tamburini, B.A., and Kedl, R.M. (2013). T cell responses: naïve to memory and everything in between. Adv. Physiol. Educ. *37*, 273–283.

Perez, O.D., and Nolan, G.P. (2002). Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry. Nat. Biotechnol. *20*, 155–162.

Perfetto, L., Briganti, L., Calderone, A., Perpetuini, A.C., Iannuccelli, M., Langone, F., Licata, L., Marinkovic, M., Mattioni, A., Pavlidou, T., et al. (2016). SIGNOR: a database of causal relationships between biological entities. Nucleic Acids Res. *44*, D548-54.

Peterson, V.M., Zhang, K.X., Kumar, N., Wong, J., Li, L., Wilson, D.C., Moore, R., McClanahan, T.K., Sadekova, S., and Klappenbach, J.A. (2017). Multiplexed quantification of proteins and transcripts in single cells. Nat. Biotechnol. *35*, 936–939.

Petryszak, R., Keays, M., Tang, Y.A., Fonseca, N.A., Barrera, E., Burdett, T., Füllgrabe, A., Fuentes, A.M.-P., Jupp, S., Koskinen, S., et al. (2016). Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. Nucleic Acids Res. *44*, D746–D752.

Potente, M., and Mäkinen, T. (2017). Vascular heterogeneity and specialization in development

and disease. Nat. Rev. Mol. Cell Biol. *18*, 477–494.

Qin, J.Y., Zhang, L., Clift, K.L., Hulur, I., Xiang, A.P., Ren, B.-Z., and Lahn, B.T. (2010). Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. PLoS One *5*, e10611.

Qiu, P., Simonds, E.F., Bendall, S.C., Gibbs, K.D., Bruggner, R. V, Linderman, M.D., Sachs, K., Nolan, G.P., and Plevritis, S.K. (2011). Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat. Biotechnol. *29*, 886–891.

Rahman, M.T., Nakayama, K., Rahman, M., Katagiri, H., Katagiri, A., Ishibashi, T., Ishikawa, M., Sato, E., Iida, K., Nakayama, N., et al. (2013). KRAS and MAPK1 gene amplification in type II ovarian carcinomas. Int. J. Mol. Sci. *14*, 13748–13762.

Rajakulendran, T., Sahmi, M., Lefrançois, M., Sicheri, F., and Therrien, M. (2009). A dimerization-dependent mechanism drives RAF catalytic activation. Nature *461*, 542–545.

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. *30*, 777–782.

Rapsomaniki, M.A., Lun, X.-K., Woerner, S., Laumanns, M., Bodenmiller, B., and Martínez, M.R. (2018). CellCycleTRACER accounts for cell cycle and volume in mass cytometry data. Nat. Commun. *9*, 632.

Rawlings, J.S., Rosler, K.M., and Harrison, D.A. (2004). The JAK/STAT signaling pathway. J. Cell Sci. *117*, 1281–1283.

Redell, M.S., Ruiz, M.J., Gerbing, R.B., Alonzo, T.A., Lange, B.J., Tweardy, D.J., Meshinchi, S., Children's Oncology Group, Z., Zhong, Z., Darnell, J., et al. (2013). FACS analysis of Stat3/5 signaling reveals sensitivity to G-CSF and IL-6 as a significant prognostic factor in pediatric AML: a Children's Oncology Group report. Blood *121*, 1083–1093.

Regot, S., Hughey, J.J., Bajar, B.T., Carrasco, S., and Covert, M.W. (2014). High-sensitivity measurements of multiple kinase activities in live single cells. Cell *157*, 1724–1734.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M., and Sabeti, P.C. (2011). Detecting novel associations in large data sets. Science *334*, 1518–1524.

Riley, N.M., and Coon, J.J. (2016). Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. Anal. Chem. *88*, 74–94.

Roberts, P.J., and Der, C.J. (2007). Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. Oncogene *26*, 3291–3310.

Romano, D., Nguyen, L.K., Matallanas, D., Halasz, M., Doherty, C., Kholodenko, B.N., and Kolch, W. (2014). Protein interaction switches coordinate Raf-1 and MST2/Hippo signalling. Nat. Cell Biol. *16*, 673–684.

Ronen, J., and Akalin, A. (2018). netSmooth: Network-smoothing based imputation for single cell RNA-seq. F1000Research *7*, 8.

Ryu, H., Chung, M., Dobrzyński, M., Fey, D., Blum, Y., Lee, S.S., Peter, M., Kholodenko, B.N., Jeon, N.L., Pertz, O., et al. (2015). Frequency modulation of ERK activation dynamics rewires cell fate. Mol. Syst. Biol. *11*, 838.

Sacco, F., Perfetto, L., Castagnoli, L., and Cesareni, G. (2012a). The human phosphatase interactome: An intricate family portrait. FEBS Lett. *586*, 2732–2739.

Sacco, F., Gherardini, P.F., Paoluzi, S., Saez-Rodriguez, J., Helmer-Citterich, M., Ragnini-Wilson, A., Castagnoli, L., and Cesareni, G. (2012b). Mapping the human phosphatome on growth pathways. Mol. Syst. Biol. *8*, 603.

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A., and Nolan, G.P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. Science *308*, 523–529.

Saez-Rodriguez, J., Alexopoulos, L.G., Zhang, M., Morris, M.K., Lauffenburger, D.A., and Sorger, P.K. (2011). Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. Cancer Res. *71*, 5400–5411.

Salgia, R. (2017). MET in Lung Cancer: Biomarker Selection Based on Scientific Rationale. Mol. Cancer Ther. *16*, 555–565.

Santarius, T., Shipley, J., Brewer, D., Stratton, M.R., and Cooper, C.S. (2010). A census of amplified and overexpressed human cancer genes. Nat. Rev. Cancer *10*, 59–64.

Santos, S.D.M., Verveer, P.J., and Bastiaens, P.I.H. (2007). Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate The mitogen-activated

protein kinase (MAPK) network is a conserved signalling module that regulates cell fate by transducing a myriad of growth-factor signals. Nat. Cell Biol. *9*.

Schapiro, D., Jackson, H.W., Raghuraman, S., Fischer, J.R., Zanotelli, V.R.T., Schulz, D., Giesen, C., Catena, R., Varga, Z., and Bodenmiller, B. (2017). histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. Nat. Methods *14*, 873–876.

Schlessinger, J. (2000). Cell Signaling by Receptor Tyrosine Kinases. Cell *103*, 211–225.

Schubert, W., Bonnekoh, B., Pommer, A.J., Philipsen, L., Böckelmann, R., Malykh, Y., Gollnick, H., Friedenberger, M., Bode, M., and Dress, A.W.M. (2006). Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. Nat. Biotechnol. *24*, 1270–1278.

Schulz, D., Zanotelli, V.R.T., Fischer, J.R., Schapiro, D., Engler, S., Lun, X.-K., Jackson, H.W., and Bodenmiller, B. (2017). Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples by Mass Cytometry. Cell Syst. *0*.

Schwamborn, K., and Caprioli, R.M. (2010). MALDI Imaging Mass Spectrometry – Painting Molecular Pictures. Mol. Oncol. *4*, 529–538.

Scialdone, A., Natarajan, K.N., Saraiva, L.R., Proserpio, V., Teichmann, S.A., Stegle, O., Marioni, J.C., and Buettner, F. (2015). Computational assignment of cell-cycle stage from single-cell transcriptome data. Methods *85*, 54–61.

Seidel, P., Roth, M., Ge, Q., Merfort, I., S'ng, C.T., Ammit, A.J., Kay, A., Ying, S., Meng, Q., Zeibecoglou, K., et al. (2011). IκBα glutathionylation and reduced histone H3 phosphorylation inhibit eotaxin and RANTES. Eur. Respir. J. *38*, 1444–1452.

Seong, H.-A., Jung, H., Ichijo, H., and Ha, H. (2010). Reciprocal negative regulation of PDK1 and ASK1 signaling by direct interaction and phosphorylation. J. Biol. Chem. *285*, 2397–2414.

Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., Choi, K., Bendall, S., Friedman, N., and Pe'er, D. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat. Biotechnol. *34*, 637–645.

Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature *546*, 431–435.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363–369.

Shi, Q., Qin, L., Wei, W., Geng, F., Fan, R., Shik Shin, Y., Guo, D., Hood, L., Mischel, P.S., and Heath, J.R. (2012). Single-cell proteomic chip for profiling intracellular signaling pathways in single tumor cells. Proc. Natl. Acad. Sci. *109*, 419–424.

Shin, I., Kim, S., Song, H., Kim, H.-R.C., and Moon, A. (2005). H-Ras-specific activation of Rac-MKK3/6-p38 pathway: its critical role in invasion and migration of breast epithelial cells. J. Biol. Chem. *280*, 14675–14683.

Silvera, D., Formenti, S.C., and Schneider, R.J. (2010). Translational control in cancer. Nat. Rev. Cancer *10*, 254–266.

Soltwisch, J., Kettling, H., Vens-Cappell, S., Wiegelmann, M., Müthing, J., and Dreisewerd, K. (2015). Mass spectrometry imaging with laser-induced postionization. Science *348*, 211–215.

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82.

Stallaert, W., Sabet, O., Bruggemann, Y., Baak, L., and Bastiaens, P. (2017). Contact inhibitory Eph signalling decouples EGFR activity from vesicular recycling to generate contextual plasticity. bioRxiv 202705.

Steelman, L., and Chappell, W. (2011). Roles of the Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR pathways in controlling growth and sensitivity to therapy-implications for cancer and aging. Aging (Albany … *3*, 192–222.

Stenmark, H. (2009). Rab GTPases as coordinators of vesicle traffic. Nat. Rev. Mol. Cell Biol. *10*, 513–525.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods *14*, 865–868.

Sugiura, Y., Honda, K., and Suematsu, M. (2015). Development of an Imaging Mass Spectrometry Technique for Visualizing Localized Cellular Signaling Mediators in Tissues. Mass

Spectrom. (Tokyo, Japan) *4*, A0040.

Sundqvist, A., Zieba, A., Vasilaki, E., Herrera Hidalgo, C., Söderberg, O., Koinuma, D., Miyazono, K., Heldin, C.-H., Landegren, U., Ten Dijke, P., et al. (2013). Specific interactions between Smad proteins and AP-1 components determine TGFβ-induced breast cancer cell invasion. Oncogene *32*, 3606–3615.

Sutherland, C. (2011). What Are the bona fide GSK3 Substrates? Int. J. Alzheimers. Dis. *2011*, 505607.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. *45*, D362–D368.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676.

Takahashi-Yanaga, F., Shiraishi, F., Hirata, M., Miwa, Y., Morimoto, S., and Sasaguri, T. (2004). Glycogen synthase kinase-3β is tyrosine-phosphorylated by MEK1 in human skin fibroblasts. Biochem. Biophys. Res. Commun. *316*, 411–415.

Takebe, N., Harris, P.J., Warren, R.Q., and Ivy, S.P. (2011). Targeting cancer stem cells by inhibiting Wnt, Notch and Hedgehog pathways. Nat. Rev. Clin. Oncol. *8*, 97–106.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

Tebbutt, N., Pedersen, M.W., and Johns, T.G. (2013). Targeting the ERBB family in cancer: couples therapy. Nat. Rev. Cancer *13*, 663–673.

Tewari, M., Hu, P.J., Ahn, J.S., Ayivi-Guedehoussou, N., Vidalain, P.-O., Li, S., Milstein, S., Armstrong, C.M., Boxem, M., Butler, M.D., et al. (2004). Systematic Interactome Mapping and Genetic Perturbation Analysis of a C. elegans TGF-β Signaling Network. Mol. Cell *13*, 469–482.

Tokarski, C., Cren-Olivé, C., Fillet, M., and Rolando, C. (2006). High-sensitivity staining of proteins for one- and two-dimensional gel electrophoresis using post migration covalent staining with a ruthenium fluorophore. Electrophoresis *27*, 1407–1416.

Türei, D., Korcsmáros, T., and Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat. Methods *13*, 966–967.

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F., et al. (2017). A pathology atlas of the human cancer transcriptome. Science *357*, eaan2507.

Valtorta, E., Misale, S., Sartore-Bianchi, A., Nagtegaal, I.D., Paraf, F., Lauricella, C., Dimartino, V., Hobor, S., Jacobs, B., Ercolani, C., et al. (2013). KRAS gene amplification in colorectal cancer and impact on response to EGFR-targeted therapy. Int. J. Cancer *133*, 1259–1265.

De Vriendt, V., De Roock, W., Di Narzo, A.F., Tian, S., Biesmans, B., Jacobs, B., Budinska, E., Sagaert, X., Rossi, S., D'Ario, G., et al. (2013). DUSP 4 expression identifies a subset of colorectal cancer tumors that differ in MAPK activation, regardless of the genotype. Biomarkers *18*, 516–524.

Wang, M., Herrmann, C.J., Simonovic, M., Szklarczyk, D., and von Mering, C. (2015). Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. Proteomics *15*, 3163–3168.

Wang, Y., Woehrstein, J.B., Donoghue, N., Dai, M., Avendaño, M.S., Schackmann, R.C.J., Zoeller, J.J., Wang, S.S.H., Tillberg, P.W., Park, D., et al. (2017). Rapid Sequential in Situ Multiplexing with DNA Exchange Imaging in Neuronal Cells and Tissues. Nano Lett. *17*, 6131–6139.

Ward, J.H. (1963). Hierarchical Grouping to Optimize an Objective Function. J. Am. Stat. Assoc. *58*, 236–244.

Waters, J.C. (2009). Accuracy and precision in quantitative fluorescence microscopy. J. Cell Biol. *185*, 1135–1148.

Wei, W., Shin, Y.S., Xue, M., Matsutani, T., Masui, K., Yang, H., Ikegami, S., Gu, Y., Herrmann, K., Johnson, D., et al. (2016). Single-Cell Phosphoproteomics Resolves Adaptive Signaling Dynamics and Informs Targeted Combination Therapy in Glioblastoma. Cancer Cell *29*, 563–573.

Wei, Z.Z., Yu, S.P., Lee, J.H., Chen, D., Taylor, T.M., Deveau, T.C., Yu, A.C.H., and Wei, L. (2014). Regulatory role of the JNK-STAT1/3 signaling in neuronal differentiation of cultured

mouse embryonic stem cells. Cell. Mol. Neurobiol. *34*, 881–893.

Wilson, N.K., Kent, D.G., Buettner, F., Shehata, M., Macaulay, I.C., Calero-Nieto, F.J., Sánchez Castillo, M., Oedekoven, C.A., Diamanti, E., Schulte, R., et al. (2015). Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell *16*, 712–724.

Wippich, F., Bodenmiller, B., Trajkovska, M.G., Wanka, S., Aebersold, R., and Pelkmans, L. (2013). Dual Specificity Kinase DYRK3 Couples Stress Granule Condensation/Dissolution to mTORC1 Signaling. Cell *152*, 791–805.

Wolf-Yadlin, A., Kumar, N., Zhang, Y., Hautaniemi, S., Zaman, M., Kim, H.-D., Grantcharova, V., Lauffenburger, D.A., and White, F.M. (2006). Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. Mol. Syst. Biol. *2*, 54.

Woyach, J.A., Johnson, A.J., and Byrd, J.C. (2012). The B-cell receptor signaling pathway as a therapeutic target in CLL. Blood *120*, 1175–1184.

Xu, Y., Li, N., Xiang, R., and Sun, P. (2014). Emerging roles of the p38 MAPK and PI3K/AKT/mTOR pathways in oncogene-induced senescence. Trends Biochem. Sci. *39*, 268–276.

Yang, X., Boehm, J.S., Yang, X., Salehi-Ashtiani, K., Hao, T., Shen, Y., Lubonja, R., Thomas, S.R., Alkan, O., Bhimdi, T., et al. (2011). A public genome-scale lentiviral expression library of human ORFs. Nat. Methods *8*, 659–661.

Yu, F.-X., Zhao, B., and Guan, K.-L. (2015). Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. Cell *163*, 811–828.

Yu, Y., Anjum, R., Kubota, K., Rush, J., Villen, J., and Gygi, S.P. (2009). A site-specific, multiplexed kinase activity assay using stable-isotope dilution and high-resolution mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. *106*, 11606–11611.

Yuan, Y., Li, C.-T., and Windram, O. (2011). Directed Partial Correlation: Inferring Large-Scale Gene Regulatory Network through Induced Topology Disruptions. PLoS One *6*, e16835.

Zadra, G., Batista, J.L., and Loda, M. (2015). Dissecting the Dual Role of AMPK in Cancer: From Experimental to Human Studies. Mol. Cancer Res. *13*, 1059–1072.

Zhang, Y., Wolf-Yadlin, A., Ross, P.L., Pappin, D.J., Rush, J., Lauffenburger, D.A., and White, F.M. (2005). Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. Mol. Cell. Proteomics *4*, 1240–1250.

Zhou, Y., Rideout, W.M., Zi, T., Bressel, A., Reddypalli, S., Rancourt, R., Woo, J.-K., Horner, J.W., Chin, L., Chiu, M.I., et al. (2010). Chimeric mouse tumor models reveal differences in pathway activation between ERBB family- and KRAS-dependent lung adenocarcinomas. Nat. Biotechnol. *28*, 71–78.

Zi, Z., Feng, Z., Chapnick, D. a, Dahl, M., Deng, D., Klipp, E., Moustakas, A., and Liu, X. (2011). Quantitative analysis of transient and sustained transforming growth factor-β signaling dynamics. Mol. Syst. Biol. *7*, 492.

Zivanovic, N., Jacobs, A., and Bodenmiller, B. (2013). A Practical Guide to Multiplexed Mass Cytometry. (Springer, Berlin, Heidelberg), pp. 95–109.

Zunder, E.R., Finck, R., Behbehani, G.K., Amir, E.-A.D., Krishnaswamy, S., Gonzalez, V.D., Lorang, C.G., Bjornson, Z., Spitzer, M.H., Bodenmiller, B., et al. (2015). Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. Nat. Protoc. *10*, 316–333.

## Acknowledgement

It has been amazing four years since I came to Zurich and joined the Bodenmiller lab. No doubt, it was the time with the deepest learning curve, for the science and research, and also for my life outside the lab. So many memories with the fantastic people around!

First of all, I would like to thank Bernd for bringing me in such a great team, for being patiently nurturing me over the last four years that I now am able to make a very big step towards my career goal. Thank you for the trust that you let me work on such a promising and important project. Most importantly, thank you for the understanding whenever I had a down time in research, or I was emotional about different things. It has been an unforgettable experience working with you, and I will be so proud to tell people that I did my PhD with you.

I also feel so grateful to know all members of the Bodenmiller lab. It is an incredible feeling that I can work together with my best friends every day. So many thanks to Vito, my teacher for computational biology and programming, thanks for always being helpful and for the great contributions to our project. And I am realizing we share so many common thoughts. Always enjoy discussing with you, about everything! And my bro James, it was so much fun working with you in the lab, chatting in the office and working out in the gym. No doubt we had extremely successful collaborations that resulted in several papers. But most importantly, thanks for teaching me all the gym and skiing skills and motivating me to be active also outside the lab. These really mean a lot to me, and to the rest of my life. Stéphane, you always joke about everything. And now I feel like I am a little bit like you, and people don't trust what I am talking about anymore. Don't know if this is good or bad. But one thing I am pretty sure about is I learned the lab techniques from the best, which made things so much easier for my entire PhD. And thanks for going running with me and torturing yourself to be so slow. And Marco, we spent so much time together in the cell culture lab, fighting for space everywhere. Trash talk with you was one of my biggest motivations to come to the lab. Of course, thanks for not being mad when I was making fun of everything about you. And to Nevena and Johanna, you are the people that have the most overlap with me in the lab. It is so grateful to think about how much helps I got form you technically and mentally. And I always enjoyed talking to you. Hart and Daniel, you are like big brothers, teaching me what are the "good science", and all of what I need for surviving. Stefanie and Andrea, you are amazingly helpful in so many things related or unrelated to work. And especially, thanks for tolerating my disorganizing. Raúl and Denis, it will always be so much fun sitting together with you guys, talking about some random things in the

lab, and about the world. Raza and Nicholas, you came to the lab with extremely strong background and expertise and it has been such a pleasant working and discussing together with you guys. And to the future of the Bodenmiller lab, Jana, Laura, Sunny, Jonas, so many great moments to share with you guys, despite our short overlap in the lab. Thanks for keeping the fantastic dynamics of the lab. Franziska and Werner, without you, my work and life would be such a mess and I probably wouldn't even get paid. So of course, thanks a lot for making things so organized that I could be more focused during my PhD. Nadine and Miquel, you were the best students! Thanks for your accompanies in work, and for being motivated on our projects. I am so sure that both of you will get plenty of successes no matter what you will do for the career, and I hope you feel those experiences working together will be helpful.

I want to thank Prof. Lucas Pelkmans and Prof. Fabian Theis. It is amazing to have you in my PhD committee. I enjoyed so much every time we talked about my projects and those interesting questions. And thanks for sharing your thoughts with me. They not only pushed my projects to move forward, but also opened my mind and promoted my personal development. I also want to say thanks to all my external collaborators during the last four years, especially Damian Szklarczyk, Attila Gábor, Prof. Christian von Mering, and Prof. Julio Saez-Rodriguez, who worked closely and intensively with me for the kinome and phosphatome project. It was such a fruitful collaboration. I felt so comfortable every time talking to you, and also I learned so much from you. I hope our project will end up somewhere that we are all happy about.

Last but most importantly, I have so many thanks to my family. Mom and dad, you inspired me when I was a little kid that I always had the dream to become a scientist. I was so proud listening to stories about the lives saved from your hands. And I hope you will be also proud of what I am doing as a biologist. Finally, Xueyang, so much to say to you, but I really want to thank you for always being understanding and supportive. There are ups and downs our lives, but no matter what happens, I always find the peacefulness with you, and I am so grateful for that.

# Xiaokang LUN

Email: xiaokang.lun@uzh.ch
Phone: +41 78 725 9839

## EDUCATION

**University of Zurich** — **Zurich, Switzerland**
PhD in Molecular Life Sciences — 05/2018

**University of Amsterdam** — **Amsterdam, the Netherlands**
MSc in Biomedical Sciences (***cum laude***) — 07/2013

**Wuhan University** — **Wuhan, China**
BSc in Biology — 06/2011

## RESEARCH EXPERIENCE

**Bernd Bodenmiller Lab** — **Zurich, Switzerland**
**University of Zurich** — 01/2014 –
PhD thesis project:
Analysis of protein abundance-dependent modulations on intercellular signaling network states and dynamics using mass cytometry

**Ana Martin-Villalba Lab** — **Heidelberg, Germany**
**German Cancer Research Center (DKFZ)** — 09/2012 – 06/2013
Master research project:
Identification of pancreatic exocrine progenitor cells with 3D organoid system and *in vivo* clonal analysis

**Coen Paulusma Lab** — **Amsterdam, the Netherlands**
**University of Amsterdam** — 12/2011 – 05/2012
Master research project:
ATP8B1 regulates intracellular vesicular trafficking in intestinal metabolism and innate immune responses

## PUBLICATIONS

[1] **X.-K. Lun**, D. Szklarczyk, A. Gábor, N. Dobberstein, V. R. T. Zanotelli, J. Saez-Rodriguez, C. von Mering, and B. Bodenmiller, "Analysis of the Human Kinome and Phosphatome Reveals Diseased Signaling Networks Induced by Overexpression," ***bioRxiv***, May 2018.

[2] M. A. Rapsomaniki*, **X.-K. Lun***, S. Woerner, M. Laumanns, B. Bodenmiller[#], and M. R. Martínez[#], "CellCycleTRACER Accounts for Cell Cycle and Volume in Mass Cytometry Data," ***Nat. Commun.***, Feb. 2018.

[3] D. Schulz, V. R. T. Zanotelli, J. R. Fischer, D. Schapiro, S. Engler, **X.-K. Lun**, H. W. Jackson, and B. Bodenmiller, "Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples by Mass Cytometry," ***Cell Syst.***, Dec. 2017.

[4] **X.-K. Lun**, V. R. T. Zanotelli, J. D. Wade, D. Schapiro, M. Tognetti, N. Dobberstein, and B. Bodenmiller, "Influence of Node Abundance on Signaling Network State and Dynamics Analyzed by Mass Cytometry," ***Nat. Biotechnol.***, Jan. 2017.

[5] D. Wollny*, S. Zhao*, I. Everlien, **X. Lun**, J. Brunken, D. Brüne, F. Ziebell, I. Tabansky, W. Weichert, A. Marciniak-Czochra, and A. Martin-Villalba, "Single-Cell Analysis Uncovers Clonal Acinar Cell Heterogeneity in the Adult Pancreas," ***Dev. Cell***, Nov. 2016.